ЭНЦИКЛОПЕДИЯ ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

II

ЭНЦИКЛОПЕДИЯ ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

II

АЛГЕБРА

АКАДЕМИЯ ПЕДАГОГИЧЕСКИХ НАУК РСФСР

ЭНЦИКЛОПЕДИЯ ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

ПОД РЕДАКЦИЕЙ
П. С. АЛЕКСАНДРОВА,
А. И. МАРКУШЕВИЧА и А. Я. ХИНЧИНА

КНИГА ВТОРАЯ

АЛГЕБРА

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ТЕХНИКО-ТЕОРЕТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1951 ЛЕНИНГРАД

Редактор *А. З. Рывкин.* Техн. редактор *Н. Я. Мурашова.*

Подписано к печати 11/XII 1950 г. Бумага 60×92¹/₁₆. 13,25 бум. л. 26,5 печ. л. 29,63 уч.-изд. л. 44 700 тип. знак. в печ. л. Т-09189. Тираж 50 000 экз. Цена книги 10 р. 40 к. Переплет 2 р. Заказ № 886.

2-я типография «Печатный Двор» им. А. М. Горького Главполиграфиздата при Совете Министров СССР. Ленинград, Гатчинская, 26.

ОГЛАВЛЕНИЕ

Предисловие 6

ВЕКТОРНЫЕ ПРОСТРАНСТВА И ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ

(*А. И. Узков*)

Глава I. **Определители и решение линейных уравнений** 11
§ 1. Векторы на плоскости 11
§ 2. Числовые векторы. Определители любого порядка 18
§ 3. Свойства определителя, вытекающие из его определения . . 21
§ 4. Перестановки. Выражение определителя порядка n 24
§ 5. Дальнейшие свойства определителя 29
§ 6. Разложение определителя по элементам ряда. Вычисление определителей 33
§ 7. Решение систем уравнений 38

Глава II. **Векторные пространства и исследование систем линейных уравнений** 42
§ 8. Векторные пространства. Абстрактная точка зрения 42
§ 9. Простейшие свойства операций над векторами 45
§ 10. Линейная зависимость векторов 49
§ 11. Подпространства 56
§ 12. Применение к системам уравнений 59
§ 13. Базис пространства. Координаты 62
§ 14. Ранг произвольной системы векторов 66
§ 15. Решение произвольных систем линейных уравнений 70
§ 16. Геометрическая интерпретация. Системы с тремя неизвестными 73
§ 17. Применение к системам уравнений высших степеней 78
§ 18. Дополнительные замечания 81

Глава III. **Линейные преобразования плоскости и трёхмерного пространства** 84
§ 19. Метрика. Скалярное произведение векторов 84
§ 20. Преобразование координат 88
§ 21. Операции над матрицами 92
§ 22. Линейные преобразования 100
§ 23. Представление линейных преобразований матрицами 105
§ 24. Геометрические свойства линейных преобразований и свойства представляющих их матриц 110
§ 25. Симметрические преобразования. Случай плоскости 114
§ 26. Симметрические преобразования трёхмерного пространства 117
§ 27. Представление произвольного линейного преобразования произведением ортогонального и симметрического 122
§ 28. Упрощение уравнений линий и поверхностей второго порядка 124
Литература 126

КОЛЬЦО МНОГОЧЛЕНОВ И ПОЛЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ

(*Л. Я. Окунев*)

Глава I. **Кольцо многочленов от одного неизвестного** 129
§ 1. Кольцо многочленов 129
§ 2. Свойства делимости многочленов от одного неизвестного . 142
§ 3. Деление на линейный двучлен $x - a$. Корни многочленов . . 159
§ 4. Многочлены над полем рациональных чисел 168
§ 5. Разложение многочленов на неприводимые множители над полем рациональных чисел. Признак неприводимости 174
§ 6. Основная теорема алгебры 188
§ 7. Проблема решения уравнений в радикалах. Двучленные уравнения 202
§ 8. Уравнения второй и третьей степеней 205
§ 9. Уравнение четвёртой степени 220
§ 10. Алгебраическое расширение и другая постановка проблемы решения уравнений в радикалах 225

Глава II. **Кольцо многочленов от нескольких неизвестных и поле рациональных функций** 235
§ 11. Кольцо многочленов от нескольких неизвестных 235
§ 12. Поле алгебраических дробей 244
§ 13. Симметрические многочлены 254
§ 14. Некоторые приложения теории симметрических многочленов 261

Глава III. **О решении алгебраических уравнений в радикалах** . . 269
§ 15. Подстановки 269
§ 16. О неразрешимости уравнений выше четвёртой степени в радикалах 273
§ 17. Группа алгебраического уравнения 281
§ 18. Уравнения с симметрической группой 294
§ 19. О разрешимости алгебраических уравнений в квадратных радикалах 300
§ 20. О разрешимости в квадратных радикалах уравнений 3-й и 4-й степеней 304
Литература 310

ЧИСЛЕННЫЕ И ГРАФИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ

(*А. П. Доморяд*)

Введение 313

Глава I. **Решение алгебраических уравнений** 317
§ 1. Постановка задачи 317
§ 2. Определение границ действительных корней 318
§ 3. Отделение корней 324
§ 4. Способ Горнера 332
§ 5. Способ Лагранжа 336
§ 6. Способ Лобачевского 343
Задачи к главе I 355

Глава II. **Решение трансцендентных уравнений** 357
§ 7. Способ линейного интерполирования и способ Ньютона . . . 357
§ 8. Обобщение способа Ньютона 363
§ 9. Способ итерации 367
§ 10. Различные способы извлечения корней из чисел 372
Задачи к главе II 378

Глава III. **Решение систем уравнений** 380
§ 11. Способ Ньютона 380
§ 12. Способ итерации 383
§ 13. Замечания о вычислении мнимых корней алгебраических уравнений 390
Задачи к главе III 391
Глава IV. **Графические методы** 392
§ 14. Уравнения с одним неизвестным 392
§ 15. Решение уравнений с помощью номограмм 399
§ 16. Решение систем уравнений 405
Задачи к главе IV 410
Добавления 412
1. Краткие исторические сведения 412
2. Советы преподавателям и рекомендуемая литература 415
Алфавитный указатель 418

ПРЕДИСЛОВИЕ

Школьный курс алгебры представляет собой своеобразное соединение сведений из различных отделов математики. Сюда входят: обобщение понятия числа (последовательное построение системы рациональных, действительных и, наконец, комплексных чисел), отнесённое нами к арифметике (см. статью И. В. Проскурякова в первой книге); изучение кольца многочленов и поля рациональных функций (охватывающее так называемые тождественные преобразования рациональных выражений) и решение алгебраических уравнений в простейших случаях, т. е. собственно алгебраический материал, отнесённый к настоящей книге; сведения о некоторых элементарных неалгебраических функциях — степенной, показательной, логарифмической, о пределах, последовательностях и простейшем ряде (геометрическая прогрессия), т. е. материал из области анализа (см. третью книгу настоящего издания), и, наконец, элементы комбинаторики, отнесённые нами в шестую книгу, где читатель найдёт также и основные сведения из теории вероятностей. Таким образом, читатель, заинтересованный научными основами школьного курса алгебры, должен знать, что он найдёт эти основы не в одной, а в нескольких книгах «Энциклопедии элементарной математики» и именно в книгах первой, второй, третьей и шестой, озаглавленных «Арифметика», «Алгебра», «Анализ» и «Разные вопросы».

Настоящая книга состоит из трёх статей. Статья А. И. Узкова даёт изложение основ того раздела математики (так называемой линейной алгебры), который вырос из теории систем алгебраических уравнений первой степени (линейных уравнений). Раздел этот (включающий, в частности, теорию определителей) освещает с единой и общей точки зрения ряд разрозненных фактов школьного курса и, кроме того, приводит к такому обобщению и углублению некоторых геометрических понятий (вектор, пространство, движение и др.), которое уже успело завоевать себе широкую область приложений.

Статья Л. Я. Окунева излагает теорию многочленов от одного и многих неизвестных и вопросы решения алгебраических уравнений в радикалах. В частности, здесь рассматривается важный для эле-

ментарной математики вопрос об условиях разрешимости алгебраических уравнений в квадратных радикалах.

В статье А. П. Доморяда, строго говоря, к алгебре относится лишь первая глава, включающая общий способ Н. И. Лобачевского для решения алгебраического уравнения любой степени с численными коэффициентами. В целом же статья представляет весьма полную сводку важнейших методов численного и графического решения алгебраических и трансцендентных уравнений, иллюстрированную конкретными примерами.

Исторические сведения по развитию теории алгебраических уравнений и других разделов алгебры не входят в эту книгу; они отнесены к «Очерку истории математики», помещаемому в седьмой книге.

Редакция

А. И. УЗКОВ

ВЕКТОРНЫЕ ПРОСТРАНСТВА И ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ

ГЛАВА I

ОПРЕДЕЛИТЕЛИ И РЕШЕНИЕ ЛИНЕЙНЫХ УРАВНЕНИЙ

§ 1. Векторы на плоскости

Под вектором в элементарной геометрии понимают направленный отрезок. Вектор обычно изображается на чертеже отрезком со стрелкой, указывающей его направление. Обозначать векторы мы будем, как правило, одной буквой жирного шрифта. Однако иногда будет применяться также и обозначение вектора двумя буквами, указывающими его начало и конец (со стрелкой сверху).

Два вектора мы будем называть *равными,* если они могут быть *совмещены параллельным перемещением.* Очевидно, что так определённое равенство векторов обладает обычными свойствами равенства: каждый вектор равен самому себе; если один вектор равен другому, то и второй равен первому; наконец, два вектора, порознь равные третьему, равны и между собою.

При работе с векторами оказывается полезным следующее соглашение об операциях над ними: под *умножением вектора на число* понимают образование нового вектора, длина которого равна длине данного вектора, умноженной на абсолютную величину данного числа, а направление либо совпадает с направлением данного вектора (если число положительно), либо противоположно этому направлению (если число, на которое умножаем, отрицательно). Кроме умножения вектора на число, определяют также сложение векторов. Его достаточно определить для двух векторов с общим началом. В таком случае *суммой* двух векторов $\overrightarrow{AB}$ и $\overrightarrow{AC}$ называют вектор, являющийся диагональю $\overrightarrow{AD}$ параллелограмма, построенного на данных векторах (рис. 1).

Эти операции позволяют составлять из данных векторов выражения вида $k_1\boldsymbol{a}_1 + k_2\boldsymbol{a}_2 + \ldots + k_n\boldsymbol{a}_n$ с любыми (действительными) числовыми коэффициентами. Такие выражения называются *линейными комбинациями* данных векторов.

Мы будем называть векторами также «отрезки» нулевой длины, т. е. такие, у которых начало и конец совпадают. Все такие

«нулевые векторы» оказываются равными между собой в смысле установленного выше определения. Нулевой вектор считается параллельным любому вектору.

Операции над векторами, определённые выше, обладают многими свойствами действий над числами: сумма не зависит от порядка слагаемых и обладает свойством ассоциативности, т. е. $(a+b)+c=a+(b+c)$; если мы имеем сумму нескольких произведений векторов на числа, то одинаковые множители можно выносить за скобки (рис. 2) и т. д. Операция, обратная сложению (вычитание), также всегда выполнима: чтобы из вектора a вычесть вектор b, достаточно образовать сумму $a+(-1)\cdot b$. Эти основные свойства операций над векторами позволяют, как это делается в элементарной алгебре, *производить формальные преобразования равенств* (переносить члены из одной части в другую, умножать обе части равенства на одно и то же число или прибавлять к ним один и тот же вектор; можно также складывать отдельно левые и правые части векторных равенств, получая при этом также справедливые равенства).

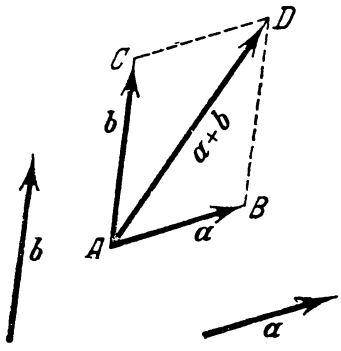

Рис. 1.

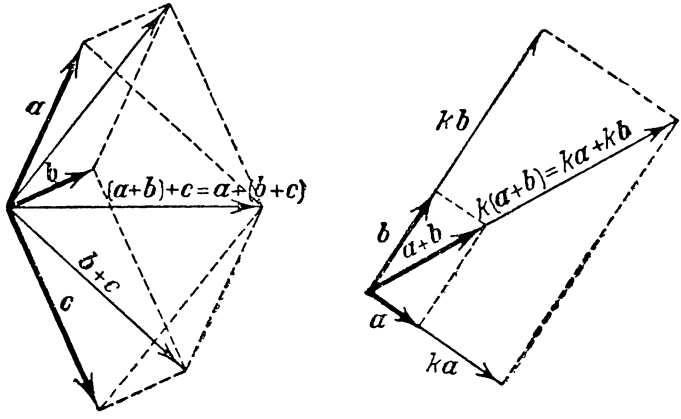

Рис. 2.

Если два вектора a и b параллельны одной и той же прямой и $a\neq 0$ (нулевому вектору), то вектор b всегда можно представить в виде $b=ka$, где k — число. Вектор, не параллельный вектору a, в этом виде представить нельзя, как это сразу следует из определения произведения вектора на число.

Мы ограничимся пока рассмотрением векторов, лежащих в одной плоскости. В этом случае сделанное выше замечание позволяет *любой вектор выразить в виде линейной комбинации двух задан-*

ных векторов, не параллельных между собою: в самом деле, если $\boldsymbol{a}$ и $\boldsymbol{b}$ — данные векторы, а $\boldsymbol{x}$ — любой вектор той же плоскости, то можно, прежде всего, параллельным перенесением совместить начала всех трёх векторов (рис. 3), затем через конец вектора $\boldsymbol{x}$ провести прямые, параллельные векторам $\boldsymbol{b}$ и $\boldsymbol{a}$, до пересечения с прямыми, на которых лежат векторы $\boldsymbol{a}$ и $\boldsymbol{b}$. Тогда из чертежа видно, что $\boldsymbol{x} = \overrightarrow{OX_1} + \overrightarrow{OY_1}$, а так как векторы $\overrightarrow{OX_1}$ и $\overrightarrow{OY_1}$ параллельны, соответственно, векторам $\boldsymbol{a}$ и $\boldsymbol{b}$, то можно так подобрать *числовые множители* x и y, чтобы было $x\boldsymbol{a} = \overrightarrow{OX_1}$ и $y\boldsymbol{b} = \overrightarrow{OY_1}$. Подставляя эти выражения в предыдущее равенство, получим: $\boldsymbol{x} = x\boldsymbol{a} + y\boldsymbol{b}$, т. е. выражение вектора $\boldsymbol{x}$ в виде линейной комбинации векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.

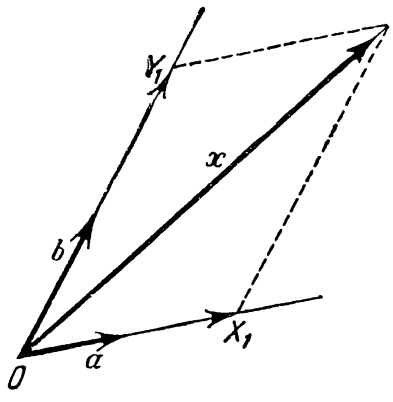

Рис. 3.

При этом замечательно, что *один и тот же вектор не может быть выражен двумя различными линейными комбинациями данных векторов*: если бы было $\boldsymbol{x} = x\boldsymbol{a} + y\boldsymbol{b} = x'\boldsymbol{a} + y'\boldsymbol{b}$, то равенство $(x' - x)\,\boldsymbol{a} = (y' - y)\,\boldsymbol{b}$ было бы также справедливым, вопреки тому, что векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не параллельны между собой.

В сказанном по существу заключена идея метода координат, известного из аналитической геометрии: если на плоскости заданы два вектора $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, то возможность однозначного представления любого вектора $\boldsymbol{x}$ в виде линейной комбинации данных, т. е. в виде $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2$, позволяет каждому вектору $\boldsymbol{x}$ поставить в соответствие два числа x_1 и x_2, которые сами однозначно определяют данный вектор. Эти числа называются *координатами* вектора $\boldsymbol{x}$ по отношению к паре векторов $\boldsymbol{e}_1$, $\boldsymbol{e}_2$. Иногда эту пару векторов называют *базисом* (или *координатной системой*) на нашей плоскости. Из указанной выше однозначности представления вектора в виде линейной комбинации непараллельных векторов следует, что *векторы равны тогда и только тогда, когда равны их координаты.* Для дальнейшего будет удобно координаты вектора $\boldsymbol{x}$ записывать в виде столбца $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

Если заданы два вектора $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2$ и $\boldsymbol{y} = y_1\boldsymbol{e}_1 + y_2\boldsymbol{e}_2$, то в силу упомянутых выше свойств операций над векторами будет также справедливо равенство $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1)\,\boldsymbol{e}_1 + (x_2 + y_2)\,\boldsymbol{e}_2$, означающее, что *координаты суммы двух векторов равны суммам соответствующих координат слагаемых.* Аналогично получается также правило, связанное с умножением вектора на число: *при*

умножении вектора на число его координаты умножаются на то же число.

Мы располагаем теперь уже достаточными знаниями, чтобы попробовать применить их к исследованию конкретного вопроса.

В качестве объекта такого применения мы выбираем хорошо известное из школьного курса исследование системы двух уравнений первой степени с двумя неизвестными, т. е. системы уравнений вида

$$\left.\begin{aligned} a_1x+b_1y&=c_1,\\ a_2x+b_2y&=c_2,\end{aligned}\right\} \tag{1}$$

где коэффициенты a_1, b_1, c_1, a_2, b_2, c_2 — данные числа, которые мы пока будем предполагать действительными.

Задача решения системы уравнений (1) состоит в определении таких значений «неизвестных» x и y, которые при подстановке их в уравнения (1) обращали бы последние в верные числовые равенства. Из школьного курса известно, что иногда система (1) решается однозначно, т. е. существует только одна система значений x и y, «удовлетворяющая» обоим уравнениям системы, иногда таких решений вовсе нет, а иногда их бесконечное множество.

То, что сказано выше о векторах, позволяет нам все эти случаи видеть совершенно непосредственно. В самом деле, рассмотрим на плоскости некоторую координатную систему и три вектора: вектор $\boldsymbol{a}$ с координатами a_1 и a_2, вектор $\boldsymbol{b}$ с координатами b_1 и b_2 и, наконец, вектор $\boldsymbol{c}$ с координатами c_1 и c_2. Если пока считать x и y известными числами, то левые части уравнений (1) будут координатами вектора $\boldsymbol{ax}+\boldsymbol{by}$. А так как эти координаты равны координатам вектора $\boldsymbol{c}$, то вектор $\boldsymbol{ax}+\boldsymbol{by}$ должен быть равен вектору $\boldsymbol{c}$. Наоборот, если нам удалось каким-либо образом подобрать такие числа x и y, чтобы выполнялось равенство $\boldsymbol{ax}+\boldsymbol{by}=\boldsymbol{c}$, то эти числа будут решением системы (1).

Таким образом, решение системы (1) совершенно эквивалентно решению одного векторного уравнения

$$\boldsymbol{ax}+\boldsymbol{by}=\boldsymbol{c}, \tag{2}$$

т. е. отысканию представления вектора $\boldsymbol{c}$ в виде линейной комбинации данных векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.

Геометрическая картина сразу подсказывает нам те возможности, которые могут здесь представиться, а именно:

1. Если оказывается, что векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не параллельны, то каждый вектор может быть представлен линейной комбинацией этих векторов, и такое представление однозначно [т. е. найдётся только одна пара значений x и y, удовлетворяющая уравнению (2)]. Это означает, что *заданная система имеет в этом случае одно решение, каковы бы ни были свободные члены* c_1 *и* c_2.

2. Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ параллельны, то решение может существовать *только* в том случае, *когда вектор* $\boldsymbol{c}$ *параллелен векторам* $\boldsymbol{a}$ и $\boldsymbol{b}$; в противном случае нужных чисел x и y найти нельзя.

3. Если все три вектора $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ параллельны, причём хотя бы один из векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ отличен от нулевого, то все решения получаются следующим образом (дальше для определённости считается, что не равен нулю вектор $\boldsymbol{a}$): придаём неизвестному y произвольное значение и переносим вектор $\boldsymbol{by}$ в правую часть: $\boldsymbol{ax} = \boldsymbol{c} - \boldsymbol{by}$. Так как вектор $\boldsymbol{c} - \boldsymbol{by}$ параллелен вектору $\boldsymbol{a}$, то оказывается возможным подобрать так значение x, чтобы было выполнено последнее равенство.

Оставшийся неразобранным случай, когда оба вектора $\boldsymbol{a}$ и $\boldsymbol{b}$ равны нулю, совсем не составляет трудностей: решение не может существовать, если вектор $\boldsymbol{c}$ отличен от нуля. Если же вектор $\boldsymbol{c}$ также равен нулю, то решениями данной системы будут служить любые пары чисел x и y.

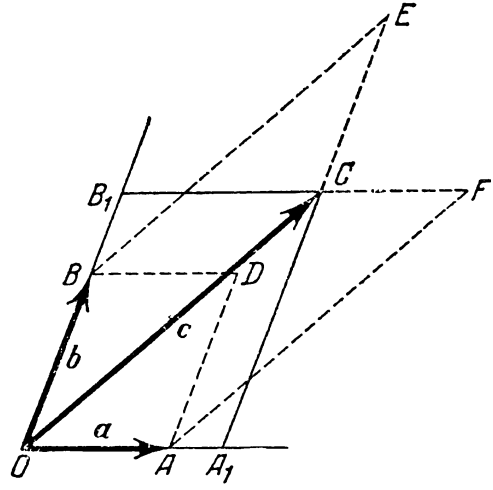

Рис. 4.

С помощью изложенных соображений можно даже получить формулы для решения системы (1) в случае, когда векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ не параллельны. В самом деле, из рис. 4 видно, что значения x и y, удовлетворяющие системе, равны, соответственно, отношениям отрезков $\frac{OA_1}{OA}$ и $\frac{OB_1}{OB}$.

Первое из этих отношений, как видно из того же рисунка, равно отношению высот параллелограммов $OCEB$ и $OADB$, у которых основанием считается вектор $\boldsymbol{b}$. Но в силу того, что основание параллелограммов — общее, отношение высот равно отношению площадей, т. е.

$$x = \frac{\text{пл } OCEB}{\text{пл } OADB}. \tag{3}$$

Аналогично, отношение $\frac{OB_1}{OB}$ равно отношению площадей параллелограммов $OCFA$ и $OBDA$, а значит,

$$y = \frac{\text{пл } OCFB}{\text{пл } OBDA}. \tag{3'}$$

Теперь не составило бы труда вычислить площади этих фигур, рассматривая подразделение их на треугольники, и тем самым полу-

чить явные формулы, выражающие значения неизвестных через коэффициенты данных уравнений. Однако в нашем случае это делается гораздо проще с помощью обычной процедуры: умножая обе части первого из уравнений данной системы (1) на b_2, а обе части второго — на $-b_1$, а затем складывая, получим для определения x уравнение

$$(a_1b_2 - a_2b_1)\,x = c_1b_2 - c_2b_1. \tag{4}$$

Точно так же, для определения y получается уравнение

$$(a_1b_2 - a_2b_1)\,y = a_1c_2 - a_2c_1. \tag{4'}$$

Легко заметить сходство этих формул с формулами (3), (3'): если образовать выражения для x и y, то у них окажется общий знаменатель, как и в формулах (3) и (3'); кроме того, этот знаменатель зависит только от коэффициентов a_1, a_2, b_1, b_2, являющихся координатами векторов $\boldsymbol{a}$ и $\boldsymbol{b}$, а в формулах (3) и (3') знаменателем является площадь параллелограмма, построенного на этих векторах. Это наталкивает на мысль выяснить геометрическое значение выражения $a_1b_2 - a_2b_1$. Выберем на плоскости базис, состоящий из двух взаимно перпендикулярных векторов $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, длина каждого из которых равна единице (рис. 5), и построим векторы $\boldsymbol{a} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2$ и $\boldsymbol{b} = b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2$. Тогда длины отрезков OA_1, OB_1, OA_2 и OB_2, взятые с надлежащими знаками, будут равны, соответственно, числам a_1, b_1, a_2 и b_2 (на чертеже все они предположены положительными). Тогда площадь параллелограмма $OACB$ будет, очевидно, выражаться так:

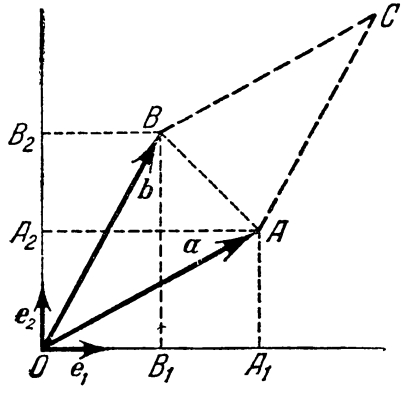

Рис. 5.

$$S_{OACB} = 2S_{OAB} = 2\,[S_{OB_1B} + S_{B_1A_1AB} - S_{OA_1A}] =$$
$$= 2\left[\frac{b_1b_2}{2} + \frac{a_2 + b_2}{2}(a_1 - b_1) - \frac{a_1a_2}{2}\right] = a_1b_2 - a_2b_1.$$

Итак, оказывается, что эта площадь в точности равна интересующему нас выражению.

Правда, во всех наших геометрических построениях имеется одна неточность: площадь в элементарно-геометрическом смысле есть величина положительная; выражение же $a_1b_2 - a_2b_1$ может быть и отрицательным. Мы не обратили внимания на это обстоятельство раньше только потому, что на наших чертежах значения всех интересующих нас величин получались положительными. Этот недостаток можно устранить, приписав площади параллелограмма опреде-

лённый знак: площадь параллелограмма, построенного на векторах $\boldsymbol{a}$ и $\boldsymbol{b}$, обычно считается положительной, если при обходе контура параллелограмма, начиная с его стороны $\boldsymbol{a}$ (в её направлении), обход совершается в ту же сторону, что и при обходе параллелограмма, посгроенного на векторах $\boldsymbol{e}_1$ и $\boldsymbol{e}_2$, начиная со стороны $\boldsymbol{e}_1$. Сделав несколько чертежей, аналогичных рис. 5, читатель без труда убедится, что определённый так знак площади всегда совпадает со знаком выражения $a_1b_2 - a_2b_1$.

Заметим, что идея снабжать площади фигур определённым знаком, в зависимости от направления обхода, применяется не только к параллелограммам, но и оказывается полезной в целом ряде вопросов, позволяя формулировать результаты наиболее общим и окончательным образом.

Выражение вида $a_1b_2 - a_2b_1$ называется *определителем второго порядка* и обозначается так:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Этим обозначением подчёркивается, что наше выражение является *функцией столбцов*, состоящих из координат векторов $\boldsymbol{a}$ и $\boldsymbol{b}$. Обращение этого выражения в нуль, как легко усмотреть из предыдущего, указывает на то, что векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ параллельны. Если $a_1b_2 - a_2b_1 \neq 0$, то решение системы (1) даётся формулами

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}; \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \tag{5}$$

вытекающими из равенств (4) и (4′). Из выясненного нами геометрического смысла определителя вытекает, что эти формулы выражают в точности то же самое, что и формулы (3) и (3′).

Для дальнейшего нам понадобятся некоторые свойства определителей. При этом будет удобнее пользоваться ещё более сокращёнными обозначениями: мы будем обозначать определитель $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ просто так: $|\boldsymbol{a}, \boldsymbol{b}|$, явно рассматривая его *как функцию двух векторов*.

А) *Определитель есть функция, линейная по каждому аргументу.*

При этом под линейностью понимают наличие двух следующих свойств [1]):

[1]) Термин «линейный» связан с тем обстоятельством, что этим свойством обладает «линейная» функция $f(x) = kx$, где k — постоянное. В этом случае из правил действий над числами следует непосредственно, что $f(x+y) = k(x+y) = f(x) + f(y)$ и $f(mx) = mf(x)$, т. е. что имеют место свойства 1) и 2), формулируемые ниже.

1) *Если значение аргумента умножить на какое-либо число, то новое значение функции получается из первоначального умножением на это же число.*

2) *Если значение аргумента равно некоторой сумме, то значение функции равно сумме её значений, получаемых при значениях аргумента, равных отдельным слагаемым.*

Говоря о линейности по каждому аргументу, мы выражаем то обстоятельство, что свойства 1) и 2) имеют место в применении к любому из двух аргументов нашей функции.

Кроме линейности, определитель обладает ещё двумя свойствами:

Б) *Если два вектора, из которых составлен определитель, равны между собою, то значение определителя равно нулю.*

В) *Определитель, составленный из базисных векторов, т. е. из векторов с координатами* $\begin{pmatrix}1\\0\end{pmatrix}$ и $\begin{pmatrix}0\\1\end{pmatrix}$ (в этом порядке), *равен единице.*

Наличие всех этих свойств проверяется непосредственным подсчётом. Их можно также получить геометрически. Для примера ограничимся алгебраической проверкой второй половины свойства А:

$$\begin{vmatrix} a_1 & b_1'+b_1'' \\ a_2 & b_2'+b_2'' \end{vmatrix} = \begin{vmatrix} a_1 & b_1' \\ a_2 & b_2' \end{vmatrix} + \begin{vmatrix} a_1 & b_1'' \\ a_2 & b_2'' \end{vmatrix} = a_1(b_2'+b_2'') - a_2(b_1'+b_1'').$$

§ 2. Числовые векторы. Определители любого порядка

Теперь естественно посмотреть, как можно применить сказанное в предыдущем параграфе к решению и исследованию систем большего числа уравнений первой степени с бóльшим числом неизвестных. Достаточно написать такую систему хотя бы с тремя неизвестными, чтобы усмотреть, что коэффициенты уравнений в этом случае группируются в столбцы, состоящие из трёх (или большего числа) чисел. Это делает невозможным в случае, если столбцы состоят более, чем из трёх чисел истолковать их как столбцы координат вектора в обычном геометрическом смысле, причём такая трудность усугубляется ещё и тем, что иногда приходится рассматривать системы уравнений с комплексными коэффициентами: ведь невозможно представить себе вектор, координаты которого комплексны.

Все эти затруднения, однако, можно обойти следующим путём, идея которого оказывается очень плодотворной и часто применяется в самых различных математических дисциплинах: элементарное понятие вектора можно обобщить так, что указанные выше трудности исчезнут сами собой, и в то же время существенные свойства векторов будут сохранены. К этому нужно добавить ещё и то, что сама геометрическая терминология будет подсказывать нам эти свойства, когда мы их будем разыскивать.

Конечно, все возникающие на этом пути понятия должны быть точно определены, т. е. сведены к известным уже нам математическим понятиям.

Прежде всего о числах. Просмотр «алгебраической» части § 1 убеждает нас в том, что во всех проведённых там алгебраических вычислениях природа рассматриваемых чисел безразлична: важно только то, что над этими числами можно производить четыре основных алгебраических действия, и эти действия подчиняются обычным законам. Это позволяет нам рассматривать каждый раз не все числа, имеющиеся в нашем распоряжении, а любые их совокупности, внутри которых можно выполнять указанные основные операции. Таким образом, мы приходим к понятию числового поля.

Числовым полем мы будем называть любую совокупность чисел, обладающую тем свойством, что сумма, разность, произведение и частное (при рассмотрении частного предполагается, что делитель отличен от нуля) *любых двух чисел этой совокупности являются числами той же совокупности* [1]).

Совокупность всех комплексных чисел удовлетворяет этому условию и поэтому является числовым полем. Точно так же числовыми полями являются совокупность всех действительных чисел и совокупность всех рациональных чисел. Эти три числовых поля наиболее часто встречаются в приложениях и поэтому наиболее важны. Однако существуют и другие поля: например, как читатель легко убедится, совокупность всех чисел вида $a + b\sqrt{2}$, где a и b — любые рациональные числа, также является полем.

Для нас в дальнейшем будет во многих случаях безразлично, какое именно числовое поле рассматривается. В таких случаях мы для удобства будем обозначать всё это поле одной буквой.

Итак, пусть K — некоторое числовое поле, а n — некоторое натуральное число.

n-мерным числовым вектором над полем K мы будем называть *любой столбец, составленный из n чисел нашего поля.*

Как мы видели, векторам на плоскости соответствуют столбцы из двух действительных чисел. Теперь в смысле только что введённого определения *сами эти столбцы* являются двухмерными числовыми векторами над полем действительных чисел.

В определённый таким образом оборот речи целесообразно ввести следующее упрощение: если рассматриваются векторы всё время над одним и тем же полем, то указание этого поля мы будем просто опускать. Точно так же будет опускаться и слово «числовые», так как никаких других векторов мы пока не будем рассматривать.

Назвав столбец из чисел вектором, естественно назвать *сами числа,* из которых столбец составлен, *координатами этого вектора.*

---

[1]) См. Э. э. м., кн. 1, И. В. Проскуряков, Понятия множества, группы, кольца и поля. Теоретические основы арифметики.

После этого уже можно определить основные операции над векторами, пользуясь аналогией с рассмотренными в § 1 «геометрическими» операциями.

Суммой двух n-мерных векторов мы назовём n-мерный вектор, координаты которого равны суммам соответствующих координат слагаемых.

Аналогично этому, *произведением n-мерного вектора на число k* (из поля K) *назовём n-мерный вектор, координаты которого равны соответствующим координатам данного вектора, умноженным на число k.*

Эти определения могут быть выражены формулами следующим образом:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}; \quad k \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_n \end{pmatrix}.$$

В дальнейшем часто окажется удобным обозначать векторы одной буквой жирного шрифта, как это делалось в § 1; в таких случаях координаты вектора будут обозначаться той же буквой светлого шрифта с индексом, указывающим номер этих координат.

Следующим определением мы придадим всей получающейся теории в некотором смысле ещё более геометрический характер: *n-мерным числовым пространством над полем K называется совокупность всех n-мерных числовых векторов над этим полем.*

Среди n-мерных числовых векторов мы особо выделим векторы

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Любой вектор $\boldsymbol{x}$ однозначно представляется линейной комбинацией этих векторов:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n.$$

Таким образом, векторы $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ играют роль б а з и с а в нашем «пространстве». Более точно мы определим смысл термина базис несколько позднее, когда нам в большей степени понадобятся его свойства.

Исходя из введённых определений, можно было бы развивать геометрию нашего пространства. Она в сильной степени напоминает обычную аналитическую геометрию. Некоторыми вопросами её мы ещё будем иметь случай заняться. Сейчас нашей целью является выяснение возможности обобщить понятие определителя так, чтобы можно было написать явные формулы, дающие решение системы уравнений с любым числом неизвестных.

После того, что было сказано об определителях второго порядка, естественно обобщение проводить так, чтобы были сохранены, конечно, с естественными изменениями, обнаруженные свойства определителей второго порядка. Необходимость изменений вызывается тем, что вместо двумерных векторов мы имеем дело с n-мерными.

Это приводит к такому определению.

Определителем порядка n называется функция

$$|a_1, a_2, \ldots, a_n|$$

от n n-мерных числовых векторов, обладающая следующими свойствами:

А) *она линейна по отношению к каждому аргументу*;

Б) *если значения двух её аргументов равны, то значение функции обращается в нуль*;

В) $|e_1, e_2, \ldots, e_n| = 1$.

Таким образом, для того чтобы сохранить при обобщении нужные нам свойства определителей второго порядка, мы их просто кладём в основу этого обобщения.

Ясно, что сформулированное определение само по себе не даёт уверенности ни в том, что такая функция существует, ни в том, что она только одна. Это может быть доказано только при дальнейшем исследовании и будет проведено в общем виде в § 5.

§ 3. Свойства определителя, вытекающие из его определения

Предположим сначала, что функция $|a_1, a_2, \ldots, a_n|$, обладающая свойствами А), Б), В), существует. Из этих свойств вытекает ряд других, которые, как мы увидим, позволяют написать явное выражение этой функции.

Рассмотрим подробнее свойство А) (линейность). В применении к первому аргументу оно запишется так:

$$\left.\begin{aligned} |\alpha a_1, a_2, \ldots, a_n| &= \alpha |a_1, a_2, \ldots, a_n|, \\ |a_1' + a_1'', a_2, \ldots, a_n| &= |a_1', a_2, \ldots, a_n| + |a_1'', a_2, \ldots, a_n| \end{aligned}\right\} \quad (1)$$

(ср. сказанное о линейности на стр. 17—18). Подобные же формулы можно написать, поставив сумму двух векторов на место не первого, а любого из аргументов.

Читатель легко заметит, что эти формулы по своему строению аналогичны формулам

$$\left.\begin{aligned} \alpha a_1 a_2 a_3 \ldots a_n &= \alpha (a_1 a_2 a_3 \ldots a_n), \\ (a_1' + a_1'') a_2 a_3 \ldots a_n &= a_1' a_2 a_3 \ldots a_n + a_1'' a_2 a_3 \ldots a_n, \end{aligned}\right\} \quad (1')$$

имеющим место для умножения чисел. Как известно, в случае чисел из формул (1'), справедливых по отношению к каждому множителю, вытекает известное общее правило умножения многочленных выражений. Например, если два первых множителя являются суммами, то

$$(\alpha_1' a_1' + \alpha_1'' a_1'')(\alpha_2' a_2' + \alpha_2'' a_2'') a_3 \ldots a_n = \alpha_1' \alpha_2' a_1' a_2' a_3 \ldots a_n +$$
$$+ \alpha_1' \alpha_2'' a_1' a_2'' a_3 \ldots a_n + \alpha_1'' \alpha_2' a_1'' a_2' a_3 \ldots a_n + \alpha_1'' \alpha_2'' a_1'' a_2'' a_3 \ldots a_n.$$

Из формул (1) также вытекает, что с определителем можно поступать, как со своеобразным произведением: *если значения одного или нескольких аргументов представляют собою суммы, то можно применять обычное правило раскрытия скобок и выносить числовые множители за знак определителя.*

Известное читателю доказательство этого для обычных произведений принимает в случае определителей такой вид (в качестве примера рассматривается случай, когда два первых аргумента являются суммами двух слагаемых, а числовые множители равны единице и в записи отсутствуют):

$$|a_1' + a_1'', \ a_2' + a_2'', \ a_3, \ldots, a_n| =$$
$$= |a_1', \ a_2' + a_2'', \ a_3, \ldots, a_n| + |a_1'', \ a_2' + a_2'', \ a_3, \ldots, a_n| =$$
$$= |a_1', \ a_2', \ a_3, \ldots, a_n| + |a_1', \ a_2'', \ a_3, \ldots, a_n| +$$
$$+ |a_1'', \ a_2', \ a_3, \ldots, a_n| + |a_1'', \ a_2'', \ a_3, \ldots, a_n|.$$

Таково первое следствие из свойств А), Б), В). Следующим важным следствием является такое:

От перемены мест двух векторов, входящих в определитель, его значение меняет знак.

В самом деле, пусть мы хотим переменить местами два первых вектора a_1 и a_2. Рассмотрим определитель

$$|a_1 + a_2, \ a_1 + a_2, \ a_3, \ldots, a_n|.$$

Он равен нулю в силу свойства Б), так как в его выражение входят два одинаковых вектора. С другой стороны, применяя к нему только что доказанное свойство, его можно представить в виде суммы четырёх определителей:

$$|a_1, \ a_1, \ a_3, \ldots, a_n| + |a_1, \ a_2, \ a_3, \ldots, a_n| +$$
$$+ |a_2, \ a_1, a_3, \ldots, a_n| + |a_2, \ a_2, \ a_3, \ldots, a_n|.$$

Из этих четырёх определителей первый и четвёртый равны нулю по той же причине, что и выше. Таким образом, из сказанного получается, что

$$|a_1, a_2, a_3, \ldots, a_n| + |a_2, a_1, a_3, \ldots, a_n| = 0,$$

а это как раз и есть доказываемое свойство.

Проведённое рассуждение в действительности является совершенно общим и не зависит от того, какие именно векторы мы переставляем. Мы ограничились рассмотрением первых двух векторов только для удобства записи.

Доказанные свойства позволяют, как уже было отмечено, найти явное выражение определителя. Чтобы вначале не затруднять читателя осложняющими деталями, связанными с общностью, мы покажем идею дальнейшего исследования, возвратившись снова к определителям второго порядка.

Пусть даны произвольные двумерные числовые векторы:

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2.$$

Посмотрим, что можно сказать о выражении $|\boldsymbol{a}, \boldsymbol{b}|$. Записывая это выражение в виде

$$|\boldsymbol{a}, \boldsymbol{b}| = |a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2, \ b_1\boldsymbol{e}_1 + b_2\boldsymbol{e}_2|$$

и применяя доказанное выше правило «раскрытия скобок», получим:

$$|\boldsymbol{a}, \boldsymbol{b}| = |a_1\boldsymbol{e}_1, \ b_1\boldsymbol{e}_1| + |a_1\boldsymbol{e}_1, \ b_2\boldsymbol{e}_2| + |a_2\boldsymbol{e}_2, \ b_1\boldsymbol{e}_1| + |a_2\boldsymbol{e}_2, \ b_2\boldsymbol{e}_2|.$$

Вынося числовые множители за знак определителя, будем иметь выражение

$$|\boldsymbol{a}, \boldsymbol{b}| = a_1b_1|\boldsymbol{e}_1, \boldsymbol{e}_1| + a_1b_2|\boldsymbol{e}_1, \boldsymbol{e}_2| + a_2b_1|\boldsymbol{e}_2, \boldsymbol{e}_1| + a_2b_2|\boldsymbol{e}_2, \boldsymbol{e}_2|.$$

В стоящей справа сумме первое и последнее слагаемые равны нулю, так как в них два аргумента определителя равны. Поэтому, заметив ещё, что $|\boldsymbol{e}_2, \boldsymbol{e}_1| = -|\boldsymbol{e}_1, \boldsymbol{e}_2|$ (правило перестановки аргументов!) и $|\boldsymbol{e}_1, \boldsymbol{e}_2| = 1$, можно переписать полученное выражение так:

$$|\boldsymbol{a}, \boldsymbol{b}| = a_1b_2 - a_2b_1.$$

Но это и есть наше исходное выражение определителя второго порядка.

Читателю рекомендуется повторить тот же ход рассуждений в применении к определителю третьего порядка $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|$, в котором

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

В этом случае применение правила раскрытия скобок к выражению

$$|a, b, c| = |a_1e_1 + a_2e_2 + a_3e_3,\ b_1e_1 + b_2e_2 + b_3e_3,\ c_1e_1 + c_2e_2 + c_3e_3|$$

даёт сумму 27 слагаемых, в каждом из которых будет стоять множителем один из определителей вида

$$|e_1, e_1, e_1|,\ |e_1, e_2, e_3|,\ |e_2, e_3, e_1|,\ |e_1, e_3, e_2|$$

и т. д.; 21 из этих определителей обращаются в нуль в силу наличия равных значений аргументов, все остальные одной или несколькими перестановками аргументов легко сводятся к определителю $|e_1, e_2, e_3|$, равному единице в силу свойства В). Например,

$$|e_2, e_3, e_1| = -|e_1, e_3, e_2| = |e_1, e_2, e_3| = 1.$$

Проводя всю выкладку подробно, читатель получит выражение

$$|a, b, c| = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1. \quad (2)$$

Из проделанных выкладок следует, что в рассмотренных случаях свойства А), Б), В) приводят к однозначно определённым выражениям для определителя, т. е. что никакие выражения, кроме полученных, не могут обладать свойствами А), Б), В). Поэтому возможно только одно из двух: либо функции, обладающей свойствами А), Б), В), вообще не существует, либо существует только одна такая функция (для каждого рассматриваемого порядка).

Последнюю альтернативу при $n = 2$ и $n = 3$ легко разрешить: достаточно проверить, обладают ли найденные выражения этими свойствами. В случае определителей второго порядка это уже было сделано раньше, а для выражения (2) читатель может выполнить такую проверку сам по образцу проведённой на стр. 18. Так как проверка показывает, что все три свойства А), Б), В) для выражения (2) имеют место, то обнаруживается и существование определителя третьего порядка.

§ 4. Перестановки. Выражение определителя порядка n

Мы покажем теперь, что принятое нами определение даёт возможность однозначно установить выражение определителя любого порядка. При этом окажется применимым в точности тот же путь, который был только что указан для определителей второго и третьего порядков.

Прежде всего мы сообщим некоторые сведения о так называемых перестановках.

Если имеется некоторое конечное число каких-либо элементов (предметов), то *любое расположение этих элементов в определённом порядке следования называется перестановкой.* Число возможных перестановок зависит только от числа имеющихся элементов.

Если число предметов равно n, то число возможных их перестановок составляет $n! = 1 \cdot 2 \ldots n$.

Часто бывает, что для рассматриваемых предметов определён некоторый «нормальный» порядок следования. Так, если элементы являются целыми числами, то нормальным порядком считается расположение их по возрастанию. Аналогично этому, если рассматриваются векторы $e_1, e_2, \ldots, e_n$, то «нормальным» расположением считается как раз то, которое только что написано. Если дана некоторая перестановка этих элементов, то естественно пытаться как-то указать, насколько она отличается от нормального расположения. Это делается следующим образом: рассмотрим в такой перестановке какие-либо два элемента; их расположение может либо быть таким же, как и при нормальном расположении, либо будет противоположным нормальному расположению. В последнем случае говорят, что рассматриваемая пара элементов образует *инверсию* или беспорядок. Можно подсчитать о б щ е е ч и с л о инверсий, образуемых всевозможными парами элементов в перестановке. Это число равно нулю тогда и только тогда, когда перестановка сама является нормальным расположением. В противном случае получается число, обязательно большее нуля. Поэтому естественно это число принять в качестве меры отклонения данной перестановки от нормального расположения.

Поясним сказанное на нескольких примерах. Перестановки

$$(3, 2, 5, 4, 1), \quad (2, 5, 3, 4, 1), \quad (3, 5, 4, 1, 2)$$

являются перестановками пяти чисел 1, 2, 3, 4 и 5. В первой из них пара чисел 3 и 2 образует инверсию, во второй эта же пара чисел инверсии не образует. Читателю предоставляется самому убедиться, что общее число инверсий в каждой из этих перестановок будет, соответственно, равно 6, 6 и 7.

Перестановка называется *чётной*, если число инверсий в ней чётное. В противном случае она называется *нечётной*.

Из написанных выше перестановок первые две — чётные, а третья — нечётная.

Если дана некоторая перестановка, то, поменяв в ней местами два элемента, мы получим некоторую новую перестановку. Такая *перемена мест двух элементов* в перестановке называется *транспозицией* этих элементов. Выполняя транспозиции несколько раз последовательно, мы будем получать всё новые и новые перестановки.

Для нас является важным то обстоятельство, что *из любой данной перестановки можно получить любую другую с помощью ряда последовательных транспозиций пар элементов.*

В самом деле, для перестановок двух элементов утверждение очевидно, так как таких перестановок всего две, и каждая из них получается из другой одной транспозицией. Это обстоятельство

позволяет для доказательства нашего утверждения в общем случае использовать индукцию по числу элементов в перестановке.

Предположим, что утверждение уже доказано для перестановок из $n-1$ элементов. Пусть даны две любые перестановки из n элементов (конечно, одних и тех же!)

$$(i_1, i_2, \ldots, i_n) \quad \text{и} \quad (j_1, j_2, \ldots, j_n).$$

Требуется рядом транспозиций вторую из них перевести в первую. Прежде всего находим среди элементов $j_1, j_2, \ldots, j_n$ элемент i_1. Пусть это будет j_k. Если $j_k \neq j_1$, то производим во второй перестановке транспозицию элементов j_k и j_1. Получаем перестановку

$$(j_k, j_2, \ldots, j_{k-1}, j_1, j_{k+1}, \ldots).$$

Если сравнить её с первой из данных перестановок, то мы увидим, что её элементы, начиная со второго, образуют некоторую перестановку элементов $i_2, i_3, \ldots, i_n$. Так как число этих элементов равно $n-1$, то согласно сделанному предположению можно рядом транспозиций превратить эту перестановку в $i_2, i_3, \ldots, i_n$, а это и нужно. Исключённый из рассмотрения случай $j_k = j_1$ ещё проще, так как в нём не требуется подготовительной транспозиции.

Вторым важным для нас фактом является то, что *при транспозиции двух элементов перестановки её чётность меняется на противоположную.*

В самом деле, если транспозиция производится над соседними элементами перестановки, — это очевидно, так как может появиться или исчезнуть лишь одна инверсия между переставляемыми элементами: расположение этих элементов относительно других элементов перестановки и других элементов между собою не изменяется.

В общем случае, когда транспонируемые элементы лежат не рядом, перемена мест указанных элементов может быть получена *транспозициями соседних элементов* следующим образом: сначала меняем местами первый из данных элементов со следующим за ним, затем во вновь полученной перестановке снова меняем местами первый из данных элементов со следующим за ним и т. д. до тех пор, пока первый из данных элементов не займёт места второго из данных. После этого второй из данных элементов переставляем с предшествующими элементами несколько раз до тех пор, пока он не займёт первоначального места первого из данных элементов. Легко усмотреть, что если между данными элементами находилось m элементов перестановки, то для «перенесения» первого элемента на место второго потребуется $m+1$ транспозиция соседних элементов, а для того, чтобы после этого перенести второй элемент на место первого, нужно m транспозиций соседних элементов. Всего для выполнения нужной нам транспозиции потребовалось $2m+1$ транспозиций соседних элементов. А так как при каждой такой транспо-

зиции чётность перестановки менялась на противоположную, то в силу нечётности числа $2m+1$ чётность полученной после перемены мест указанных элементов перестановки будет противоположна чётности данной.

Применим изложенные соображения для получения выражения определителя произвольного порядка, предположив пока, что определитель [т. е. функция, удовлетворяющая условиям А), Б) и В) определения § 2] существует.

Рассмотрим, прежде всего, частный случай:

$$|e_{j_1}, e_{j_2}, \ldots, e_{j_n}|, \tag{1}$$

т. е. случай, когда определитель составлен из базисных векторов $e_1, \ldots, e_n$. Здесь ясно, что значение будет равно нулю, коль скоро хотя бы два из векторов $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ совпадают [в силу свойства Б)]. Таким образом, остаётся рассмотреть только значение определителя (1), когда все векторы e_{j_k} различны. Но тогда ряд векторов $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ является перестановкой векторов $e_1, e_2, \ldots, e_n$, а поэтому рядом последовательных транспозиций его можно обратить в ряд $e_1, e_2, \ldots, e_n$. При этом данный определитель обратится в определитель $|e_1, e_2, \ldots, e_n|$, равный 1, в силу условия В). Учитывая теперь, что при каждой транспозиции входящих в определитель векторов определитель только меняет знак (см. § 3), мы получаем следующий результат:

Определитель $|e_{j_1}, e_{j_2}, \ldots, e_{j_n}|$, *в котором все векторы* $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ *различны, равен* $+1$ *или* -1, *смотря по тому, является ли перестановка* $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ *векторов* $e_1, e_2, \ldots, e_n$ *чётной или нечётной. Если два из векторов* $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ *совпадают, то значение определителя равно нулю.*

Конечно, чётность перестановки $e_{j_1}, e_{j_2}, \ldots, e_{j_n}$ рассматриваемых векторов $e_1, e_2, \ldots, e_n$ совпадает с чётностью перестановки их индексов $j_1, j_2, \ldots, j_n$.

Пусть теперь даны произвольные векторы $a_1, a_2, \ldots, a_n$. Их координаты целесообразно обозначать ещё одним индексом, помещаемым перед номером вектора. Так, вторая координата третьего вектора будет обозначаться a_{23}, четвёртая координата первого вектора будет a_{41} и т. д. Таким образом,

$$a_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \end{pmatrix}, \quad a_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{n2} \end{pmatrix}, \quad \ldots, \quad a_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{nn} \end{pmatrix}.$$

Эти векторы, как однажды было замечено, выражаются через векторы $e_1, e_2, \ldots, e_n$ так:

$$a_1 = e_1 a_{11} + e_2 a_{21} + \ldots + e_n a_{n1},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_n = e_1 a_{1n} + e_2 a_{2n} + \ldots + e_n a_{nn}.$$

Поэтому определитель $|a_1, a_2, \ldots, a_n|$ может быть переписан следующим образом:

$$|e_1 a_{11} + \ldots + e_n a_{n1},\ e_1 a_{12} + \ldots + e_n a_{n2}, \ldots,\ e_1 a_{1n} + \ldots + e_n a_{nn}|. \quad (2)$$

Используя указанную в § 3 возможность оперирования с определителем как со своеобразным «произведением», можем переписать это выражение в виде

$$|a_1, a_2, \ldots, a_n| = \sum |e_{j_1}, e_{j_2}, \ldots, e_{j_n}|\, a_{j_1 1}\, a_{j_2 2} \ldots a_{j_n n},$$

где стоящая справа сумма распространяется на в с е комбинации индексов $j_1, j_2, \ldots, j_n$. Однако в силу сделанного выше замечания определители $|e_{j_1}, e_{j_2}, \ldots, e_{j_n}|$, в которых хотя бы два индекса $j_1, j_2, \ldots, j_n$ совпадают, равны нулю. Поэтому среди членов в правой части остаются только те, в которых все эти индексы j различны, причём стоящие там определители $|e_{j_1}, e_{j_2}, \ldots, e_{j_n}|$ равны ± 1. Поэтому мы окончательно имеем выражение

$$|a_1, a_2, \ldots, a_n| = \sum \pm a_{j_1 1}\, a_{j_2 2} \ldots a_{j_n n}, \quad (3)$$

в котором знак $+$ приписывается тем членам, у которых перестановка индексов $j_1, j_2, \ldots, j_n$ чётная, а знак $-$ приписывается членам, у которых эта перестановка нечётная. Ясно, что число членов в сумме (3) равно числу различных перестановок из n элементов, т. е. равно $n!$

То, что наши рассуждения приводят к совершенно однозначному выражению, позволяет сформулировать следующую теорему:

Если существует функция $|a_1, a_2, \ldots, a_n|$ *от* n *n-мерных числовых векторов, удовлетворяющая условиям* А), Б), В) *нашего определения, то её значения определяются выражением* (3).

Другими словами, требования А), Б), В) о д н о з н а ч н о определяют нашу функцию.

Естественно, возникает вопрос о том, насколько эта функция будет определена, если некоторые из указанных требований отбросить. Оказывается, что в таком случае однозначность уже не будет иметь места. Особенно интересен результат, который получается при отбрасывании условия В). Так как этот результат нам будет полезен в дальнейшем, мы его сформулируем и докажем.

Если дана функция $F(a_1, a_2, \ldots, a_n)$ *от* n *n-мерных числовых векторов* $a_1, a_2, \ldots, a_n$, *удовлетворяющая условиям* А) *и* Б), *то её значения выражаются формулой*

$$F(a_1, a_2, \ldots, a_n) = |a_1, a_2, \ldots, a_n| F(e_1, e_2, \ldots, e_n),$$

где $|a_1, a_2, \ldots, a_n|$ *означает определитель, составленный из векторов* $a_1, a_2, \ldots, a_n$, *т. е. выражение* (3).

Таким образом, достаточно знать только одно значение $F(e_1, \ldots, e_n)$ нашей функции, чтобы однозначно определить все её значения.

Для доказательства достаточно заметить, что в предыдущих рассуждениях мы пользовались свойством В) только в случаях, когда требовалось указать значение $|e_1, e_2, \ldots, e_n|$. Поэтому аналогия с умножением многочленов и свойство менять знак при перемене мест векторов сохраняются и для рассматриваемой функции $F(a_1, a_2, \ldots, a_n)$: достаточно в проведённых выше доказательствах заменить $|a_1, a_2, \ldots, a_n|$ на $F(a_1, a_2, \ldots, a_n)$. Но если такую замену сделать в доказательстве формулы (3), мы вместо неё получим:

$$F(a_1, a_2, \ldots, a_n) = F(e_1, e_2, \ldots, e_n) \sum \pm a_{j_1 1}\, a_{j_2 2} \ldots a_{j_n n},$$

т. е. как раз то, что требуется.

§ 5. Дальнейшие свойства определителя

Мы должны теперь изучить свойства выражения (3) предыдущего параграфа. При этом окажется, что сами свойства А), Б), В) для этого выражения имеют место. Тем самым впервые обнаружится существование определителя. Ряд других свойств позволит сравнительно просто вычислять определители любого порядка и, как увидим дальше, применить определители для решения систем уравнений с любым числом неизвестных.

Чтобы не вводить новых терминов, будем называть «определителем» само выражение (3) предыдущего параграфа и примем для него более развёрнутое обозначение: вместо $|a_1, a_2, \ldots, a_n|$ будем писать:

$$\begin{vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots & \ldots \\ a_{n1} & a_{n2} \ldots a_{nn} \end{vmatrix},$$

как это уже делалось для определителей второго и третьего порядков.

Рассмотрим более подробно выражение (3):

$$\begin{vmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \cdot & \cdot \quad \cdot \\ a_{n1} & a_{n2} \dots a_{nn} \end{vmatrix} = \sum \pm a_{j_1 1} a_{j_2 2} \dots a_{j_n n}. \qquad (*)$$

Из самой записи правой части видно, что *в каждый член* стоящей там суммы *входит ровно по одному множителю из каждого столбца*. Вспоминая теперь, что индексы $j_1, j_2, \dots, j_n$ образуют просто некоторую перестановку индексов $1, 2, \dots, n$, можем сказать также, что *каждый член суммы содержит ровно по одному множителю из каждой строки*, написанной выше квадратной таблицы. Важно также обратить внимание на то, что знак $+$ или $-$ зависит только от чётности перестановки $j_1, j_2, \dots, j_n$, т. е. *только от тех мест* в таблице, на которых стоят множители рассматриваемого члена.

Из этих наблюдений немедленно обнаруживается свойство линейности А). Если все элементы одного из столбцов нашей таблицы умножить на некоторое число k, то, поскольку все члены стоящей справа суммы содержат точно по одному элементу из рассматриваемого столбца, вся сумма просто умножится на число k, и мы получаем первое условие линейности:

От умножения всех элементов одного из столбцов определителя на число k значение определителя умножается на это число.

Другими словами, общий множитель всех элементов одного из столбцов можно вынести за знак определителя.

Представим себе теперь, что нам задан определитель

$$\begin{vmatrix} a_{11} & a_{12} \dots a'_{1k} + a''_{1k} \dots a_{1n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{n1} & a_{n2} \dots a'_{nk} + a''_{nk} \dots a_{nn} \end{vmatrix} = |a_1, a_2, \dots, a_k, \dots, a_n|.$$

Применяя к нему равенство (*), замечаем, что в каждом слагаемом полученной суммы будет содержаться один из множителей $a'_{jk} + a''_{jk}$. Раскрывая в каждом слагаемом скобки, разобьём сумму на две суммы, которые будут отличаться от исходной только тем, что в них вместо элемента k-го столбца стоят элементы одного из двух столбцов

$$a'_k = \begin{pmatrix} a'_{1k} \\ \cdot \\ \cdot \\ \cdot \\ a'_{nk} \end{pmatrix} \quad \text{или} \quad a''_k = \begin{pmatrix} a''_{1k} \\ \cdot \\ \cdot \\ \cdot \\ a''_{nk} \end{pmatrix}.$$

Таким образом, значения этих частичных сумм будут равны, соответственно, определителям

$$|a_1, a_2, \ldots, a'_k, \ldots, a_n| \text{ и } |a_1, a_2, \ldots, a''_k, \ldots, a_n|,$$

а исходный определитель равен сумме этих определителей.

Рассмотрим теперь выражение (*) в случае, когда два из столбцов $a_1, a_2, \ldots, a_n$ совпадают. Для определённости предположим, что совпадают первые два столбца (рассуждения, которые приводятся ниже, не зависят от номеров этих столбцов). Возьмём какой-либо член $\pm a_{j_1 1} a_{j_2 2} \ldots a_{j_n n}$ суммы (*). Он содержит множитель $a_{j_1 1}$ из 1-го столбца и множитель $a_{j_2 2}$ из 2-го столбца. Поменяв местами индексы j_1 и j_2, получаем новый член $\pm a_{j_2 1} a_{j_1 2} \ldots a_{j_n n}$, суммы (*). Знаки этих членов определяются перестановками $j_1, j_2, \ldots, j_n$ и $j_2, j_1, \ldots, j_n$. А так как эти перестановки получаются одна из другой транспозицией индексов j_1 и j_2, то они неизбежно различной чётности, так что знаки в обоих случаях необходимо различны. Что касается самих произведений $a_{j_1 1}, a_{j_2 2} \ldots a_{j_n n}$ и $a_{j_2 1} a_{j_1 2} \ldots a_{j_n n}$, то они совпадают, так как в силу равенства столбцов a_1 и a_2 будет: $a_{j_1 2} = a_{j_1 1}$ и $a_{j_2 2} = a_{j_2 1}$. Итак, в случае совпадения двух столбцов определителя $|a_1, a_2, \ldots, a_n|$ в его выражении каждому члену будет соответствовать такой же член с противоположным знаком, т. е. в этом случае значение определителя равно нулю. Таким образом, свойство Б) также имеет место.

Остаётся проверить свойство В). Для этого рассмотрим определитель

$$\begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{vmatrix}$$

Применяя к нему равенство (*), видим, что в сумме остаётся только один член, отличный от нуля, — это произведение стоящих по диагонали единиц. Соответствующая этому члену перестановка индексов, как легко видеть, есть $1, 2, 3, \ldots, n$. Так как эта перестановка не содержит инверсий, то она чётная, а следовательно, значение нашего определителя равно $+1$.

С этого момента мы можем утверждать, что смысл термина «определитель», который мы ввели в начале этого параграфа, совпадает с первоначальным смыслом, установленным определением § 2. Одновременно, конечно, обнаружено, что выражение (*) обладает и свойствами, доказанными в начале § 3.

К уже известным нам свойствам полезно добавить ещё два свойства, важные для вычисления определителей и вытекающие из уже доказанных:

Значение определителя не меняется, если к элементам одного из его столбцов прибавить элементы другого столбца, умноженные на одно и то же число.

При доказательстве удобно воспользоваться нашими старыми обозначениями. Пусть, например, к элементам первого столбца определителя $|a_1, a_2, a_3, \ldots, a_n|$ мы прибавим соответствующие элементы третьего столбца, умноженные на число k. Тогда получается определитель $|a_1 + ka_3, a_2, a_3, \ldots, a_n|$. Пользуясь линейностью определителя по первому аргументу, получаем:

$$|a_1 + ka_3, a_2, a_3, \ldots, a_n| = |a_1, a_2, a_3, \ldots, a_n| + k\,|a_3, a_2, a_3, \ldots, a_n|.$$

Но второе слагаемое в правой части равно нулю в силу свойства Б). Этим всё доказано.

Значение определителя не изменится, если его столбцы заменить соответствующими строками.

Операция замены столбцов соответствующими строками называется *транспонированием* таблицы определителя.

Рассмотрим определитель

$$D = \begin{vmatrix} a_{11} & a_{21} \ldots a_{n1} \\ a_{12} & a_{22} \ldots a_{n2} \\ \ldots & \ldots \\ a_{1n} & a_{2n} \ldots a_{nn} \end{vmatrix},$$

получающийся из данного транспонированием. В полученной таблице *первый индекс служит уже номером столбца, а второй — номером строки.* Поэтому, применяя для явного выражения определителя сумму (*), мы должны записать:

$$D = \sum \pm a_{1k_1}\, a_{2k_2} \ldots a_{nk_n}. \qquad (**)$$

Переставим теперь в каждом произведении $a_{1k_1}\, a_{2k_2} \ldots a_{nk_n}$ множители таким образом, чтобы в них вторые индексы шли в порядке возрастания. Конечно, при этом расположение первых индексов нарушится, и мы получим запись того же произведения в виде $a_{j_1 1}\, a_{j_2 2} \ldots a_{j_n n}$. Чётность перестановки $j_1, j_2, \ldots, j_n$ будет той же, что и чётность перестановки $k_1, k_2, \ldots, k_n$, так как при проделанной выше операции приведения вторых индексов к нормальному расположению мы должны выполнить столько же транспозиций первых индексов, сколько их было выполнено над вторыми. Таким образом, мы обнаружили, что член $\pm a_{1k_1}\, a_{2k_2} \ldots a_{nk_n}$ суммы (**) попросту совпадает с членом суммы (*), полученным перестановкой множителей. Но так как это совпадение имеет место для каждого члена, то сумма (**) равна сумме (*), что и нужно.

Доказанное свойство определителя показывает равноправность его столбцов и строк. Отсюда следует, что те свойства, кото-

рые мы выше формулировали и доказали для столбцов определителя, имеют место также и для его строк. Это относится вообще ко всем свойствам, которые присущи выражению (*), если только в формулировке их участвуют столбцы нашей квадратной таблицы целиком.

§ 6. Разложение определителя по элементам ряда. Вычисление определителей

Рассмотрим произвольный определитель

$$D=\begin{vmatrix} a_{11} \dots a_{1,k-1} & a_{1k} & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & a_{2k} & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots\dots & & \\ a_{n1} \dots a_{n,k-1} & a_{nk} & a_{n,k+1} \dots a_{nn} \end{vmatrix}$$

и выберем в нём какой-либо столбец, например k-й. Этот столбец можно представить в виде суммы n столбцов и переписать заданный определитель так:

$$D=\begin{vmatrix} a_{11} \dots a_{1,k-1} & a_{1k}+0 & +\dots+0 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0+a_{2k} & +\dots+0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots\dots & & & \\ a_{n1} \dots a_{n,k-1} & 0+0 & +\dots+a_{nk} & a_{n,k+1} \dots a_{nn} \end{vmatrix}$$

Пользуясь теперь свойством А) определителя, можно представить данный нам определитель в виде суммы n определителей, у которых в k-м столбце имеется самое большее один отличный от нуля элемент. Вынося этот элемент за знак определителя в качестве множителя, получаем представление данного определителя в виде такой суммы:

$$D=a_{1k}\begin{vmatrix} a_{11} \dots a_{1,k-1} & 1 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots\dots & & \\ a_{n1} \dots a_{n,k-1} & 0 & a_{n,k+1} \dots a_{nn} \end{vmatrix}+\dots$$

$$\dots+a_{nk}\begin{vmatrix} a_{11} \dots a_{1,k-1} & 0 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots\dots & & \\ a_{n1} \dots a_{n,k-1} & 1 & a_{n,k+1} \dots a_{nn} \end{vmatrix}. \quad (1)$$

Определители, стоящие в правой части равенства, получаются из данного определителя заменой всех элементов k-го столбца нулями, кроме одного элемента, который заменяется единицей.

Условимся называть каждый из таких определителей *алгебраическим дополнением* того элемента, который был заменён еди-

ницей. При этом условии равенство (1) позволяет высказать такое утверждение:

Определитель равен сумме произведений элементов одного из его столбцов на их алгебраические дополнения.

Этот результат позволяет свести вычисление определителей к вычислению определителей низших порядков, так как мы сейчас установим, что алгебраические дополнения с точностью до знака совпадают с некоторыми определителями более низкого порядка, просто получаемыми из данного определителя.

Назовём *минором* данного определителя, соответствующим элементу a_{jk}, определитель, получаемый из данного вычёркиванием строки и столбца, на пересечении которых лежит рассматриваемый элемент.

Вычислим сначала алгебраическое дополнение элемента определителя, стоящего в верхнем левом углу, т. е. элемента a_{11}. Это алгебраическое дополнение по определению равно определителю

$$\begin{vmatrix} 1 & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ 0 & a_{32} & \dots & a_{3n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Мы не изменим значения этого определителя, если прибавим ко второму его столбцу первый столбец, умноженный на $-a_{12}$, к третьему столбцу — первый столбец, умноженный на $-a_{13}$, и т. д. После этого получим определитель

$$\begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{32} & a_{33} & \dots & a_{3n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix}. \tag{2}$$

Отсюда видно, прежде всего, что алгебраическое дополнение элемента зависит т о л ь к о от столбцов

$$\begin{pmatrix} a_{22} \\ a_{32} \\ \cdot \\ \cdot \\ \cdot \\ a_{n2} \end{pmatrix}, \quad \begin{pmatrix} a_{23} \\ a_{33} \\ \cdot \\ \cdot \\ \cdot \\ a_{n3} \end{pmatrix}, \quad \dots, \quad \begin{pmatrix} a_{2n} \\ a_{3n} \\ \cdot \\ \cdot \\ \cdot \\ a_{nn} \end{pmatrix} \tag{3}$$

и поэтому может рассматриваться как функция этих столбцов. Сами написанные выше столбцы, конечно, являются $(n-1)$-мерными числовыми векторами.

В силу известных нам свойств определителей наш определитель (2), очевидно, будет обладать следующими свойствами:

1. Если умножить один из столбцов (3) на число k, то значение определителя (2) умножится на то же число.

2. Если один из столбцов (3) является суммой других столбцов, то определитель (3) будет равен сумме таких же определителей, полученных из данного заменой рассматриваемого столбца отдельными слагаемыми.

3. Если два из столбцов (3) совпадают, то определитель (2) обращается в нуль.

4. Если столбцы (3) равны, соответственно, столбцам

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad \ldots, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix},$$

то определитель (2) равен единице, так как он обращается в определитель, диагональными элементами которого являются единицы, а остальные элементы равны нулю.

Короче говоря, наш определитель является функцией от $n-1$ $(n-1)$-мерных числовых векторов, удовлетворяющей условиям А), Б), В), т. е. *будет определителем порядка* $n-1$, *составленным из столбцов* (3), *а этот определитель и есть минор данного определителя, соответствующий элементу* a_{11}.

Теперь нетрудно вычислить и алгебраическое дополнение любого элемента a_{jk} данного определителя. Для этого в рассматриваемом алгебраическом дополнении

$$\begin{vmatrix} a_{11} & \ldots & a_{1,k-1} & 0 & a_{1,k+1} & \ldots & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{j-1,1} & \ldots & a_{j-1,k-1} & 0 & a_{j-1,k+1} & \ldots & a_{j-1,n} \\ a_{j1} & \ldots & a_{j,k-1} & 1 & a_{j,k+1} & \ldots & a_{jn} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \ldots & a_{n,k-1} & 0 & a_{n,k+1} & \ldots & a_{nn} \end{vmatrix}$$

будем последовательно переставлять k-й столбец с предыдущими столбцами, пока он не окажется на первом месте, а затем будем аналогичным образом переставлять j-ю строку. При этом общее

число транспозиций столбцов будет $k-1$, а число транспозиций строк равно $j-1$. Таким образом, для того чтобы компенсировать изменение знака определителя при каждой транспозиции строк и столбцов, мы должны умножить полученный в конце концов определитель на $(-1)^{j+k-2}$, т. е. на $(-1)^{j+k}$. Этим способом мы находим, что рассматриваемое алгебраическое дополнение равно определителю

$$(-1)^{j+k}\begin{vmatrix} 1 & a_{j1} & \dots & a_{j,k-1} & a_{j,k+1} & \dots & a_{jn} \\ 0 & a_{11} & \dots & a_{1,k-1} & a_{1,k+1} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & a_{j-1,1} & \dots & a_{j-1,k-1} & a_{j-1,k+1} & \dots & a_{j-1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & a_{n1} & \dots & a_{n,k-1} & a_{n,k+1} & \dots & a_{nn} \end{vmatrix}.$$

Для того чтобы свести его вычисление к вычислению определителя более низкого порядка, осталось только применить к нему формулированную выше теорему. Согласно ей этот определитель равен сумме произведений элементов первого столбца на их алгебраические дополнения. При этом в сумме фактически остаётся один член, соответствующий первому элементу столбца, так как остальные его элементы равны нулю. Но алгебраическое дополнение первого элемента первого столбца, как мы видели выше, равно минору, соответствующему этому элементу. Поэтому окончательно получаем, что рассматриваемое алгебраическое дополнение равно

$$(-1)^{j+k}\begin{vmatrix} a_{11} & \dots, & a_{1,k-1} & a_{1,k+1} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{j-1,1} & \dots, & a_{j-1,k-1} & a_{j-1,k+1} & \dots & a_{j-1,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & \dots, & a_{n,k-1} & a_{n,k+1} & \dots & a_{nn} \end{vmatrix}.$$

Этот результат можно выразить совсем короткой формулой, если условиться обозначать алгебраическое дополнение элемента a_{jk} через A_{jk}, а соответствующий этому элементу минор — через M_{jk}. В таком случае наш результат запишется в виде

$$A_{jk}=(-1)^{j+k}M_{jk}.$$

Сама теорема, которая была формулирована в начале параграфа, может быть теперь записана в виде формулы

$$D=\sum_{j=1}^{n} a_{jk}A_{jk}=a_{1k}A_{1k}+a_{2k}A_{2k}+\dots+a_{nk}A_{nk}, \qquad (4)$$

или

$$D=\sum_{j=1}^{n}(-1)^{j+k}M_{jk}a_{jk}, \qquad (4')$$

где k означает номер того столбца, по элементам которого разлагается определитель.

Равенство (4') может быть использовано для вычисления определителей в любом случае. Однако наиболее легко его применять в случаях, когда в рассматриваемом столбце многие элементы равны нулю. При вычислении определителей это условие само по себе редко бывает выполненным, и поэтому бывает полезно искусственно создавать нулевые элементы в определителе, пользуясь тем, что к его столбцам можно, не изменяя значения определителя, прибавлять другие столбцы, умноженные на произвольные числа. Этим мы уже пользовались выше при вычислении алгебраического дополнения элемента a_{11}.

Следует заметить, что данное выше определение алгебраического дополнения по форме несимметрично: в нём строки и столбцы играют внешне различную роль. Однако, полученное выражение $A_{jk} = (-1)^{j+k}M_{jk}$ показывает, что алгебраическое дополнение элемента определителя на самом деле не изменяется при замене строк столбцами, так как при этом не изменяются ни сумма номеров строки и столбца, содержащих данный элемент, ни дополнительный минор. Отсюда следует, что найденные для столбцов результаты можно и в этом случае применять к строкам определителя. Например, формула (4) переписывается для строк так: $D = \sum_{k=1}^{n} a_{jk}A_{jk}$. Многочисленные примеры вычисления определителей собраны в задачнике Фаддеева и Соминского, указанном в помещённом ниже списке литературы. В этом задачнике приведены подробные указания, упрощающие вычисления в частных случаях.

В заключение приведём ещё одно свойство определителя, вытекающее из доказанной теоремы:

Сумма произведений элементов одного из столбцов определителя на алгебраические дополнения соответствующих элементов другого столбца равна нулю.

В самом деле, пусть речь идёт о k-м и l-м столбцах. Тогда в силу формулы (4) будем иметь:

$$D = \sum_{j=1}^{n} a_{jl}A_{jl}.$$

Подставляя сюда вместо элементов l-го столбца элементы k-го столбца, получим выражение

$$\sum_{j=1}^{n} a_{jk}A_{jl}. \qquad (5)$$

Так как изменение здесь заключается только в том, что элементы

l-го столбца заменены элементами k-го столбца (а элементы последнего оставлены на месте!), то выражение (5) будет равно определителю, получаемому из D путём указанной замены. Но при этом, так как вместо l-го столбца нужно подставить k-й столбец, оставляя k-й на месте, в новом определителе будут два одинаковых столбца. Это означает, что выражение (5) будет равно нулю, что и требовалось доказать. Аналогичный результат для строк даёт формулу

$$\sum_{k=1}^{n} a_{ik}A_{jk} = 0 \text{ при } i \neq j.$$

§ 7. Решение систем уравнений

Теперь у нас подготовлен вспомогательный аппарат для того, чтобы получить общие формулы, дающие решение системы n уравнений с n неизвестными.

Рассмотрим такую систему:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n. \end{aligned}\right\} \tag{1}$$

Определитель

$$|a_1, \ldots, a_n| = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} = D, \tag{2}$$

составленный из коэффициентов при неизвестных в системе (1), называется *определителем этой системы уравнений.*

Возьмём алгебраические дополнения элементов первого столбца определителя (2): $A_{11}, A_{21}, \ldots, A_{n1}$, помножим на них, соответственно, обе части каждого из уравнений системы и сложим левые и правые части всех полученных уравнений. Пользуясь тем, что сумма произведений элементов любого столбца определителя на алгебраические дополнения элементов другого столбца равна нулю, замечаем, что коэффициенты при неизвестных x_i ($i = 2, 3, \ldots, n$) в уравнении, полученном в результате указанного сложения, обращаются в нуль. Поэтому полученное уравнение будет иметь вид

$$\left(\sum_{i=1}^{n} a_{i1}A_{i1}\right) x_1 = \sum_{i=1}^{n} b_i A_{i1}. \tag{3}$$

Коэффициент при x_1 будет равен определителю системы (1). Но, замечая теперь, что правая часть уравнения (3) отличается от коэффициента левой части только тем, что вместо элементов a_{i1} в неё входят элементы b_i, видим, что правая часть также может быть записана в виде определителя: *этот определитель получается из определителя системы заменой первого столбца столбцом свободных членов.* Если полученный в результате такой замены определитель обозначить D_1, то уравнение (3) перепишется в виде

$$Dx_1 = D_1.$$

Аналогично, если умножить обе части каждого из уравнений соответственно на алгебраические дополнения элементов некоторого k-го столбца определителя D, то после сложения получим уравнение

$$Dx_k = D_k,$$

где D_k означает определитель, полученный из D заменой k-го столбца столбцом свободных членов.

В итоге, делая такие преобразования при всех $k = 1, 2, \ldots, n$, получим систему уравнений

$$Dx_k = D_k \qquad (k = 1, 2, \ldots, n). \tag{4}$$

Эта система, как показывает сам процесс её получения, *является следствием данной системы,* т. е. любое решение системы (1) будет удовлетворять также и системе (4).

Но значения неизвестных, удовлетворяющие системе уравнений (4), однозначно определяются из этой системы, если только определитель D отличен от нуля. Именно,

$$x_k = \frac{D_k}{D} \qquad (k = 1, 2, \ldots, n). \tag{5}$$

Таким образом, исходная система в этом случае может удовлетворяться *только указанными значениями неизвестных,* и мы получаем следующий промежуточный результат:

Если определитель системы n уравнений с n неизвестными отличен от нуля, то эта система не может иметь более одного решения.

Предшествующие рассуждения не могли гарантировать, что и уравнения (1) являются следствиями уравнений системы (4). В общем случае это неверно. Однако если определитель D системы (1) отличен от нуля, то можно непосредственной проверкой показать, что значения неизвестных, определяемые формулами (5), удовлетворяют исходной системе.

В самом деле, разлагая определитель D_k по элементам k-го столбца, можно записать формулы (5) в виде

$$x_k = \frac{b_1A_{1k} + b_2A_{2k} + \ldots + b_nA_{nk}}{D} = \frac{\sum_{i=1}^{n} b_iA_{ik}}{D}.$$

Подставляя теперь эти выражения вместо неизвестных в какое-нибудь j-е уравнение системы (1), получим в левой части этого уравнения выражение

$$\frac{a_{j1}\left(\sum_{i=1}^{n} b_iA_{i1}\right) + a_{j2}\left(\sum_{i=1}^{n} b_iA_{i2}\right) + \ldots + a_{jn}\left(\sum_{i=1}^{n} b_iA_{in}\right)}{D}.$$

Раскрывая теперь скобки и группируя отдельно члены, содержащие различные коэффициенты b_1, b_2, .., b_n, получим выражение

$$\frac{1}{D}\left[b_1\left(\sum_{k=1}^{n} a_{jk}A_{1k}\right) + b_2\left(\sum_{k=1}^{n} a_{jk}A_{2k}\right) + \ldots + b_n\left(\sum_{k=1}^{n} a_{jk}A_{nk}\right)\right].$$

Теперь легко видеть, что только одна, именно j-я, сумма в скобках будет отлична от нуля: она будет равна определителю системы как сумма произведений элементов j-й строки определителя на их алгебраические дополнения, остальные суммы обратятся в нуль как суммы произведений элементов j-й строки определителя на алгебраические дополнения другой строки. Эти замечания дают возможность переписать предыдущее выражение в виде

$$\frac{1}{D} b_jD,$$

что после сокращения на D даёт b_j, т. е. наше выражение равно правой части j-го уравнения системы.

Так как номер j уравнения может быть любым, мы обнаружили, что значения (5) неизвестных x_1, ..., x_n удовлетворяют всем уравнениям системы. Тем самым показано существование решения.

Полученный результат, будучи объединён со сформулированной выше теоремой единственности решения, даёт следующую теорему.

Основная теорема о системах уравнений. *Система n уравнений с n неизвестными, определитель которой отличен от нуля, всегда имеет решение. Это решение единственно и определяется следующим правилом: значение каждого из неизвестных равно дроби, знаменателем которой является определитель си-*

стемы, а числитель получается из определителя системы заменой столбца коэффициентов при искомом неизвестном столбцом свободных членов.

Приведённое правило решения системы носит название *правила Крамера.*

Полученная основная теорема ничего не говорит ни о существовании, ни о единственности решения в случае, если определитель системы равен нулю. Использованные нами соображения могут дать только условие, при котором решения заведомо не существуют (это будет, например, в случае, когда определитель D равен нулю, а хотя бы один из определителей D_k не равен нулю). Подобного рода чисто отрицательные результаты недостаточны, и поэтому, минуя их, мы сразу перейдём к подробному исследованию систем уравнений, составляющему содержание следующей главы.

ГЛАВА II

ВЕКТОРНЫЕ ПРОСТРАНСТВА И ИССЛЕДОВАНИЕ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

§ 8. Векторные пространства. Абстрактная точка зрения

Введённое в предыдущей главе понятие n-мерного числового пространства по существу является не обобщением, а только аналогом понятия векторного пространства элементарной геометрии.

Для того чтобы получаемые нами результаты можно было применять как к тому, так и к другому пространству, целесообразно отказаться от каких-либо ограничений природы тех объектов, которые мы называем векторами. Это можно сделать, исходя из такого определения:

Любая совокупность L каких-либо элементов называется векторным пространством над данным числовым полем K, если:

1. *Установлено некоторое правило, ставящее в соответствие каждым двум элементам $\boldsymbol{a}$ и $\boldsymbol{b}$ нашей совокупности некоторый третий элемент $\boldsymbol{a}+\boldsymbol{b}$ той же совокупности, называемый суммой данных элементов $\boldsymbol{a}$ и $\boldsymbol{b}$.*

2. *Установлено некоторое другое правило, ставящее в соответствие каждому элементу $\boldsymbol{a}$ нашей совокупности и каждому числу k из поля K некоторый элемент $k\boldsymbol{a}$ совокупности L.*

3. *Оба эти правила удовлетворяют следующим требованиям* (аксиомам):

I. *Для любых элементов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$ совокупности L имеют место соотношения:*

а) $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$ (коммутативность),

б) $(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$ (ассоциативность).

II. *В совокупности L существует элемент* 0 (нулевой элемент) *такой, что $\boldsymbol{a}+0=\boldsymbol{a}$ для любого элемента $\boldsymbol{a}$ из нашей совокупности.*

III. *Для каждого элемента $\boldsymbol{a}$ из L существует такой элемент $-\boldsymbol{a}$, называемый «противоположным» для $\boldsymbol{a}$, что $\boldsymbol{a}+(-\boldsymbol{a})=0$.*

IV. *Для любых элементов* a *и* b *совокупности* L *и любых чисел* k_1 *и* k_2 *поля* K *имеют место соотношения*:

а) $k_1(k_2a) = (k_1k_2)a$,

б) $(k_1 + k_2)a = k_1a + k_2a$,

в) $k_1(a + b) = k_1a + k_1b$.

V. *Для любого элемента* a *из* L *имеет место соотношение* $1a = a$, т. е. умножение на число 1 не изменяет элемента нашей совокупности L.

При таком определении естественно называть любой элемент рассматриваемой совокупности L *вектором.*

По поводу введённого определения следует заметить следующее:

Мы не случайно не указали, какие именно правила определяют сумму векторов и произведение вектора на число. Эти правила могут быть любыми, лишь бы были выполнены наложенные выше требования.

Также не случайно, что у нас на первый план выдвинулось понятие *векторного пространства*, а не понятие вектора.

Причина этого в том, что интересующие нас свойства векторов проявляются не на отдельных экземплярах этих объектов, а в поведении их, так сказать, в коллективе, во всей рассматриваемой совокупности.

Что касается употребления в формулировке аксиом знака равенства, то мы раз навсегда условимся, что равенство обозначает лишь то, что предметы, стоящие по обе его стороны, просто совпадают: это — один и тот же предмет. При таком употреблении знака равенства свойства транзитивности, симметрии и рефлексивности становятся чисто логическими его свойствами и не нуждаются в установлении особых соглашений.

Прежде чем переходить к более подробному изучению свойств векторных пространств, мы покажем на нескольких примерах, насколько большую свободу даёт нам введённое только что определение.

Легко видеть, прежде всего, что как совокупность «геометрических» векторов плоскости, так и совокупность n-мерных числовых векторов (при любом заданном n) являются векторными пространствами во введённом нашим определением смысле.

Однако теперь уже можно указать весьма большое число других примеров векторных пространств.

Пример 1. Обозначим через F_n совокупность многочленов $a_0 + a_1x + \ldots + a_nx^n$, степень каждого из которых не превосходит данного числа n, а коэффициенты берутся из рассматриваемого числового поля K. Для таких многочленов в элементарной алгебре определена операция сложения и указано правило умножения многочлена на любое число. При этом сложение и умножение на числа из поля K

не выводят нас за пределы рассматриваемой совокупности F. Кроме того, те требования, которые были сформулированы в виде аксиом I—V, в рассматриваемом случае выполнены.

Это означает, что совокупность F при указанных операциях сложения и умножения на число является векторным пространством над полем K, а сами многочлены степени не выше n можно рассматривать как векторы этого «пространства».

Отметим, что если мы ограничимся только рассмотрением многочленов, степень которых в точности равна n, то мы не получим векторного пространства: сумма таких многочленов может иметь меньшую степень, так что не будет элементом нашей совокупности.

Пример 2. *Матрицей* с m строчками и n столбцами называется *любая таблица чисел*, имеющая вид

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Если все входящие в эту таблицу числа берутся из поля K, то мы будем говорить о матрице над этим полем.

Для матриц приняты такие определения операций: *произведением матрицы на число* называется матрица, получаемая из данной умножением всех входящих в неё чисел на это число; суммой матриц называется матрица, получаемая из данных двух матриц сложением соответствующих (т. е. стоящих в них на одинаковых местах) чисел. Сумма матриц определена только тогда, когда обе рассматриваемые матрицы имеют одинаковое число строк и одинаковое число столбцов.

При этих соглашениях совокупность всех матриц с m строками и n столбцами над данным полем K будет векторным пространством над этим полем, ибо, как легко проверить, все требования сформулированного выше определения здесь опять выполнены.

Пространство m-мерных числовых векторов над полем K является частным случаем определённого только что «пространства матриц»: наши матрицы обращаются в числовые векторы, если $n = 1$, т. е. если в них имеется только один столбец.

Пример 3. Обозначим через C совокупность всех непрерывных функций $f(x)$, определённых на некотором отрезке $[a, b]$ значений x и принимающих действительные значения. В математическом анализе определяется, что значит «сложить» две данные функции или «умножить» данную функцию на действительное число. При этом оказывается, что сумма двух непрерывных на данном отрезке функций будет также непрерывна на этом отрезке, и это же относится к произведению непрерывной функции на действительное число. Кроме того, хотя это редко доказывается в явной форме, опреде-

лённые так сложение и умножение на число удовлетворяют требованиям, выраженным в аксиомах I—V.

Но сказанное означает, что совокупность C при введённых так операциях над функциями является также векторным пространством над полем действительных чисел.

Пример 4. Совокупность F_∞ всех многочленов с коэффициентами из данного поля K также является векторным пространством, если считать сумму многочленов и произведение многочлена на число определёнными обычным образом.

Этими примерами далеко не исчерпываются не только все вообще существующие векторные пространства, но даже те из них, которые оказываются наиболее важными в современной математике.

Полезность введённого в этом параграфе аксиоматического или «абстрактного» определения векторного пространства состоит именно в том, что оно позволяет изучать одновременно очень большое количество различных «конкретных» пространств. В самом деле, все результаты, которые удаётся получить, исходя только из введённого определения, необходимо будут верны в любом случае, когда условия определения выполнены.

Нужно добавить, что на самом деле связь между «абстрактными» и «конкретными» результатами здесь не только в том, что из общих теорем абстрактной теории получаются конкретные следствия, но и в том, что известные результаты в случае отдельных частных «конкретных» пространств, например совокупности векторов на плоскости или в обычном трёхмерном пространстве, *позволяют предвидеть факты, имеющие место в общем случае.* Тем самым мы получаем средство устанавливать связи иногда даже между очень отдалёнными на первый взгляд разделами математики.

Развитие математики в течение конца XIX в. и те работы, которые ведутся в настоящее время, указывают много таких связей. Их простое перечисление не может принести пользы, а более подробное выяснение существа дела завело бы нас слишком далеко от основной темы. Поэтому мы ограничимся сделанными замечаниями и перейдём к систематическому исследованию свойств векторных пространств, начиная с простейших.

§ 9. Простейшие свойства операций над векторами

Мы начнём с того, что выведем простейшие следствия из определения векторного пространства, относящиеся к операциям над векторами.

Из нашего соглашения об употреблении знака равенства и из того, что сумма двух векторов однозначно определена, вытекает следующее хорошо известное правило оперирования с равенствами:

К обеим частям векторного равенства можно прибавить один и тот же вектор, не нарушая справедливости равенства.

В самом деле, равенство $a = b$ выражает, что буквами a и b обозначен один и тот же вектор. Поскольку сумма $a + c$ однозначно определена, она будет одним и тем же вектором независимо от того, как обозначено первое слагаемое: $a + c = b + c$.

По той же причине *обе части векторного равенства можно умножать на один и тот же числовой множитель без нарушения равенства.*

Правило сложения векторов ставит в соответствие любым двум векторам a и b их сумму $a + b$. Если нам нужно сложить не два, а несколько векторов, то мы вынуждены эту операцию проводить в несколько приёмов, каждый из которых состоит в сложении двух векторов. Например, если мы имеем три вектора a, b и c, то можно представить себе такие различные комбинации при их сложении

$$(a + b) + c, \quad a + (b + c), \quad (a + c) + b, \quad \ldots,$$

где скобки, как обычно, показывают порядок выполнения операций. Законы коммутативности и ассоциативности, содержащиеся в аксиоме I, позволяют утверждать, что во всех этих случаях мы получим один и тот же результат. Действительно, для первых двух из написанных выражений это очевидно в силу самой формулировки аксиомы I, а для третьего наше утверждение следует из такой цепочки равенств:

$$(a + c) + b = b + (a + c) = (b + a) + c = (a + b) + c,$$

каждое из которых получается однократным применением равенств а) или б) аксиомы I.

Естественно посмотреть, будет ли это сохраняться и в случае сложения большего числа векторов.

Методом математической индукции может быть доказано, что имеет место следующий общий закон:

Сумма любого числа векторов не зависит от того порядка, в котором производится операция сложения данных векторов.

Этот результат даёт право при записи суммы вообще не писать скобок или расставлять скобки произвольно, если это почему-либо выгодно для дальнейшего оперирования с рассматриваемой суммой.

Доказательство сформулированного выше утверждения может быть проведено следующим образом:

Рассмотрим сначала суммы такого специального вида:

$$(\ldots(((a_1 + a_2) + a_3) + a_4) + \ldots + a_{n-1}) + a_n,$$

т. е. такие, в которых сначала складываются два первых слагаемых, затем к их сумме прибавляется третье слагаемое, затем к сумме этих слагаемых прибавляется четвёртое и т. д. Будем называть эти суммы *каноническими* и при выписывании таких сумм вовсе не будем писать скобок. Заметим,

прежде всего, что каноническую сумму n слагаемых $a_1+a_2+\ldots+a_n$ можно по самому её определению представить в следующем виде: $(a_1+\ldots+a_k)+a_{k+1}+\ldots+a_n$, т. е. как каноническую сумму $n-k+1$ слагаемых, из которых первое само является канонической суммой первых k слагаемых данной суммы.

Покажем, что *в канонической сумме можно произвольно менять порядок слагаемых*. В самом деле, для сумм из двух элементов справедливость утверждения гарантируется аксиомой I. Пусть утверждение доказано для канонических сумм, состоящих из $n-1$ элементов, и пусть дана сумма $a_1+a_2+\ldots+a_n$, для которой требуется показать, что она равна сумме $a_{i_1}+a_{i_2}+\ldots+a_{i_n}$, где $i_1, i_2, \ldots, i_n$ — какая-либо перестановка индексов 1, 2, ..., n. Представим первую и вторую суммы, соответственно, в виде $(a_1+\ldots+a_{n-1})+a_n$ и $(a_{i_1}+\ldots+a_{i_{n-1}})+a_{i_n}$. Если $i_n=n$, то видно, что дело сводится к перестановке членов в сумме $a_1+\ldots+a_{n-1}$, т. е. для этого случая всё доказано. В противном случае вектор a_{i_n} содержится среди векторов $a_1, \ldots, a_{n-1}$, причём его в сумме $n-1$ слагаемых можно переставить на последнее место, не изменяя суммы $a_1+\ldots+a_{n-1}$. Поэтому можно считать, что этот вектор уже стоит на последнем месте, т. е. что $i_n=n-1$. Векторы $a_{i_1}, \ldots, a_{i_{n-2}}, a_{i_{n-1}}$ в этом случае образуют перестановку векторов $a_1, \ldots, a_{n-2}, a_n$. Поэтому в скобках второй суммы можно (опять по предположению индукции!) переставить слагаемые так, чтобы они стояли в порядке $a_1, \ldots, a_{n-2}, a_n$. Таким образом, остаётся только установить, что равны суммы $(a_1+\ldots+a_{n-1})+a_n$ и $(a_1+\ldots+a_{n-2}+a_n)+a_{n-1}$. Этот последний шаг может быть проведён так: первая сумма по определению равна $(a_1+\ldots+a_{n-2})+a_{n-1}+a_n$, так что её можно рассматривать как сумму трёх слагаемых $(a_1+\ldots+a_{n-2}, a_{n-1}$ и $a_n)$. Поэтому скобки могут быть переставлены, и мы можем написать:

$$(a_1+\ldots+a_{n-2})+a_{n-1}+a_n=(a_1+\ldots+a_{n-2})+(a_{n-1}+a_n).$$

После этого в силу коммутативности двух слагаемых можно написать равенство

$$(a_1+\ldots+a_{n-2})+(a_{n-1}+a_n)=(a_1+\ldots+a_{n-2})+(a_n+a_{n-1}).$$

Теперь можно снова воспользоваться законом ассоциативности для трёх слагаемых и написать:

$$(a_1+\ldots+a_{n-2})+(a_n+a_{n-1})=((a_1+\ldots+a_{n-2})+a_n)+a_{n-1}.$$

Но правая часть последнего равенства есть не что иное, как сумма $a_1+\ldots+a_{n-2}+a_n+a_{n-1}$, или $(a_1+\ldots+a_{n-2}+a_n)+a_{n-1}$. Равенство доказано.

Теперь покажем, что *сумма двух канонических сумм равна канонической сумме всех их слагаемых*.

Действительно, если вторая сумма содержит лишь одно слагаемое, то утверждение следует просто из определения канонической суммы. Пусть утверждение верно в случае, если второе слагаемое содержит $n-1$ слагаемых, и пусть дана сумма $(a_1+\ldots+a_m)+(b_1+\ldots+b_n)$, в которой вторая из написанных канонических сумм содержит n слагаемых. Тогда написанную сумму можно переписать в виде $(a_1+\ldots+a_m)+((b_1+\ldots+b_{n-1})+b_n)$, т. е. рассмотреть как сумму трёх слагаемых $(a_1+\ldots+a_m)$, $(b_1+\ldots+b_{n-1})$ и b_n. В таком случае мы имеем право переставить скобки так: $((a_1+\ldots+a_m)+(b_1+\ldots+b_{n-1}))+b_n$. При этом в наружных скобках оказывается сумма двух канонических сумм, вторая из которых содержит уже $n-1$ слагаемых. В силу сделанного предположения выражение в наружных скобках может быть переписано в виде $(a_1+\ldots+a_m+b_1+\ldots+b_{n-1})$, а всё написанное выражение в виде $(a_1+\ldots+a_m+b_1+\ldots+b_{n-1})+b_n$. Но полученное выражение есть просто каноническая сумма всех слагаемых обеих сумм.

Остаётся рассмотреть произвольные суммы и показать, что *любым образом записанная сумма равна канонической сумме её слагаемых*. Это утверждение тривиально для сумм двух векторов, так как сумма двух векторов сама является канонической. Предположим, что утверждение уже доказано для сумм, состоящих из $n-1$ или меньшего числа слагаемых. Если дана сумма n слагаемых $a_1, \ldots, a_n$ с какой-либо расстановкой скобок, то её можно представить как сумму двух слагаемых (для этого нужно только усмотреть, какое из указанных сложений производится последним). Каждое из слагаемых является само суммой части слагаемых данной суммы. Поскольку число слагаемых обеих частичных сумм меньше n, каждая из них равна канонической сумме входящих в неё слагаемых. Поэтому вся рассматриваемая сумма как сумма двух канонических сумм будет канонической суммой всех её слагаемых. Этим доказательство закончено.

Подобным же образом можно рассмотреть и произведения векторов на числа. Например, *произведение* $(k_1+\ldots+k_n)(a_1+\ldots+a_m)$ *равно сумме всевозможных попарных произведений вида* $k_i a_j$ *при различных* i *и* j.

Доказывается это индукцией по числу m слагаемых векторов, а для $m=1$ — индукцией по числу n. Детали рассуждения не представляют труда, и мы оставляем его читателю.

Для большего удобства мы условимся ещё писать множители в таких произведениях в произвольном порядке, приняв по определению, что $k\boldsymbol{a}=\boldsymbol{a}k$ для любого вектора $\boldsymbol{a}$ и любого числа k из поля K. Тогда правила действий над векторами примут обычный вид, хорошо известный из элементарной алгебры. Нужно только следить, чтобы в каждом рассматриваемом произведении все множители, кроме одного, были числовыми.

Следует особо остановиться на содержании аксиом III и V.

Первая из них утверждает существование хотя бы одного элемента $\mathbf{0}$ (нулевого вектора), удовлетворяющего условию $\boldsymbol{a}+\mathbf{0}=\boldsymbol{a}$ при любом векторе $\boldsymbol{a}$. Легко, однако, видеть, что *такой вектор может быть только один*: если бы были два таких вектора $\mathbf{0}$ и $\mathbf{0}'$, то сумма $\mathbf{0}+\mathbf{0}'=\mathbf{0}'+\mathbf{0}$ равнялась бы одновременно $\mathbf{0}$ и $\mathbf{0}'$. Поэтому векторы $\mathbf{0}$ и $\mathbf{0}'$ были бы равны, вопреки первоначальному предположению, что они различны.

Аналогично этому можно установить, что *для каждого вектора* $\boldsymbol{x}$ *существует только один противоположный вектор*: если бы таких векторов было два $(-\boldsymbol{x})$ и $(-\boldsymbol{x})'$, то мы имели бы такую цепочку равенств, каждое из которых вытекает из принятых аксиом:

$$(-\boldsymbol{x})=(-\boldsymbol{x})+\mathbf{0}=(-\boldsymbol{x})+(\boldsymbol{x}+(-\boldsymbol{x})')=((-\boldsymbol{x})+\boldsymbol{x})+(-\boldsymbol{x})'=$$
$$=(\boldsymbol{x}+(-\boldsymbol{x}))+(-\boldsymbol{x})'=\mathbf{0}+(-\boldsymbol{x})'=(-\boldsymbol{x})'+\mathbf{0}=(-\boldsymbol{x})'.$$

Теперь можно легко установить, что если какой-либо вектор $\boldsymbol{x}$ удовлетворяет условию $\boldsymbol{a}+\boldsymbol{x}=\boldsymbol{a}$, хотя бы при одном векторе $\boldsymbol{a}$, то $\boldsymbol{x}=\mathbf{0}$. В самом деле, прибавляя к обеим частям первоначального равенства вектор $(-\boldsymbol{a})$, получим $\boldsymbol{x}=\mathbf{0}$.

Обнаруженное свойство нулевого вектора даёт возможность показать, что *произведение любого вектора* $\boldsymbol{a}$ на число 0 *равно нулевому вектору.*

Действительно, в силу аксиом IV и V имеем: $\boldsymbol{a}+0\cdot\boldsymbol{a}=1\cdot\boldsymbol{a}+0\cdot\boldsymbol{a}=(1+0)\cdot\boldsymbol{a}=\boldsymbol{a}$. Это означает, что вектор $0\cdot\boldsymbol{a}$ обладает указанным свойством, а значит, он равен нулевому вектору.

Сделанное замечание позволяет не опасаться путаницы от того, что число нуль и нулевой вектор мы обозначаем одинаковым образом. Их свойства аналогичны, а в каждом отдельном случае сразу видно, с числом или с вектором мы имеем дело.

Если имеется векторное равенство, одна из частей которого содержит слагаемым вектор $\boldsymbol{a}$, то, прибавляя к обеим частям противоположный вектор $-\boldsymbol{a}$, получим равенство, отличающееся от первоначального только тем, что вектор $\boldsymbol{a}$ «перенесён» из одной его части в другую с противоположным знаком. Подобного рода формальными преобразованиями равенств мы будем далее пользоваться без особых оговорок.

§ 10. Линейная зависимость векторов

Доказанные в предыдущем параграфе свойства операций над векторами позволяют действовать с ними так же свободно, как с обычными числами или с многочленами. Мы больше не будем возвращаться к этим элементарным свойствам и обратимся к понятию, играющему в дальнейшем основную роль, — понятию линейной зависимости.

Уже в § 1 мы пользовались термином «линейная комбинация» векторов. При рассмотрении любого векторного пространства L мы будем также называть выражения вида $k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\ldots+k_n\boldsymbol{a}_n$ *линейными комбинациями* векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$, если только коэффициенты $k_1, k_2, \ldots, k_n$ принадлежат числовому полю K, над которым рассматривается наше пространство. Обратим внимание читателя на то, что при этом не исключается случай, когда все коэффициенты равны нулю. Поэтому нулевой вектор всегда можно представить линейной комбинацией любых заданных векторов.

Пусть теперь M — произвольная система векторов нашего пространства. *Эта система называется линейно независимой, если линейная комбинация векторов, принадлежащих системе, может быть равна нулю только в том случае, когда все коэффициенты равны нулю.* В случае, если можно указать хотя бы одну линейную комбинацию векторов системы, коэффициенты которой не все равны нулю и которая тем не менее равна нулю, говорят, что данная система *линейно зависима.*

Легко видеть, что если на плоскости взять любые два непараллельных вектора, то они будут линейно независимы: никакая их линейная комбинация с отличными от нуля коэффициентами не может

быть равна нулю. Такое же положение будет, если в «обычном» трёхмерном пространстве взять три вектора, не параллельные одной плоскости.

Наоборот, если на плоскости взять любую систему из трёх векторов, то она уже будет линейно зависимой: в этом случае хотя бы один из трёх векторов будет линейной комбинацией двух других, например $\boldsymbol{a}=k_1\boldsymbol{b}+k_2\boldsymbol{c}$. Но, перенося все члены этого равенства в одну часть, мы получим $1\boldsymbol{a}-k_1\boldsymbol{b}-k_2\boldsymbol{c}=0$. А это означает, что нашлась линейная комбинация данных векторов, которая равна нулю, несмотря на отличие от нуля её коэффициентов (коэффициент 1 при векторе $\boldsymbol{a}$, очевидно, отличен от нуля).

Только что проведённое рассуждение позволяет доказать следующую простую, но важную теорему.

Теорема. *Система векторов, содержащая более одного элемента, линейно зависима тогда и только тогда, когда хотя бы один из её векторов представляется линейной комбинацией остальных.*

Таким образом, содержащееся в формулировке этой теоремы условие почти эквивалентно первоначальному определению линейной зависимости. Некоторым дефектом этого условия является лишь то, что оно не может быть применено к одному вектору: если система состоит из одного вектора, то говорить о его выражении через «остальные» нельзя без некоторой натяжки; в то же время говорить о линейной зависимости или линейной независимости такой системы в смысле нашего определения можно, так как можно рассматривать линейные комбинации векторов, состоящие из одного слагаемого.

Именно, возможность применять первоначальное определение без каких-либо оговорок делает его более удобным, чем принятие в качестве определения линейной зависимости условия, указываемого нашей теоремой.

Доказательство. Пусть система M линейно зависима. Тогда по определению линейной зависимости можно указать линейную комбинацию $k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\ldots+k_n\boldsymbol{a}_n$ некоторых векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ системы, которая равна нулю и в которой хотя бы один коэффициент отличен от нуля. Не ограничивая общности, можно считать, что $k_1 \neq 0$ (в противном случае можно было бы изменить нумерацию рассматриваемых векторов). Тогда из равенства $k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2+\ldots+k_n\boldsymbol{a}_n=0$ умножением обеих его частей на $k_1^{-1}=\frac{1}{k_1}$ получим: $\boldsymbol{a}_1=-\frac{k_2}{k_1}\boldsymbol{a}_2-\ldots-\frac{k_n}{k_1}\boldsymbol{a}_n$, т. е. один из векторов системы выражается линейной комбинацией остальных.

Наоборот, если какой-либо вектор $\boldsymbol{a}_1$ данной системы представляется комбинацией остальных векторов, например $\boldsymbol{a}_1=k_2\boldsymbol{a}_2+\ldots+k_n\boldsymbol{a}_n$, то перенос всех членов в одну часть приводит к

линейной комбинации, равной нулю и имеющей отличный от нуля (хотя бы) один из коэффициентов.

Из определения сразу следует, что *если система векторов содержит нулевой вектор, то она обязательно линейно зависима*: в этом случае выражение $k\mathbf{0}$ представляет собою линейную комбинацию векторов системы, содержащую лишь одно слагаемое. Эта линейная комбинация будет равна нулю при любом значении коэффициента k, в частности при $k \neq 0$.

Точно так же получается сразу, что *любая часть линейно независимой системы векторов будет линейно независимой системой.*

В самом деле, если бы эта часть была линейно зависимой, то было бы возможным составление линейной комбинации её векторов, равной нулю и содержащей хотя бы один отличный от нуля коэффициент. Но эта линейная комбинация была бы также линейной комбинацией векторов всей системы, что, очевидно, противоречит линейной независимости последней.

Условимся называть две системы векторов *эквивалентными, если каждый вектор любой из них можно выразить линейной комбинацией векторов другой системы.*

Так, если на плоскости взять три вектора x, y, z, связанные соотношением $z = x + y$, то системы векторов $\{x, y\}$, $\{x, y, z\}$, а также и $\{y, z\}$ будут эквивалентными.

Например, эквивалентность первых двух из этих систем следует из справедливости равенств

$$\begin{aligned} x &= 1x, \\ y &= 1y, \end{aligned} \qquad \begin{aligned} x &= 1x, \\ y &= 1y, \\ z &= x + y, \end{aligned}$$

содержание которых сводится именно к тому, что каждый вектор одной из этих двух систем является линейной комбинацией векторов другой системы.

Определённая так эквивалентность обладает следующими свойствами:

Любая система векторов эквивалентна самой себе.

Если одна система векторов эквивалентна другой системе, то и другая система эквивалентна первой.

Если каждая из двух данных систем векторов эквивалентна одной и той же третьей системе, то две первые системы также эквивалентны друг другу.

Убедиться в справедливости этих утверждений предоставляется самому читателю.

Важное свойство линейно независимых систем векторов указывает следующая

Теорема о замене. *Пусть конечная линейно независимая система $a_1, a_2, \ldots, a_n$ векторов такова, что каждый её вектор*

является линейной комбинацией векторов некоторой другой системы M. Тогда число векторов в системе M не может быть меньше n, причём можно произвести замену n векторов системы M векторами $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ так, чтобы полученная после этой замены система M' была эквивалентна первоначальной системе M.

Доказательство. Поскольку в формулировку теоремы входит натуральное число n — число элементов рассматриваемой конечной системы, естественно для доказательства воспользоваться индукцией.

Начнём со случая $n=1$, когда первая система состоит лишь из одного вектора. Из того, что имеет место линейная независимость, следует, что вектор $\boldsymbol{a}_1 \neq 0$. С другой стороны, также по условию теоремы, можно выразить вектор $\boldsymbol{a}_1$ линейной комбинацией некоторых векторов $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_s$ системы M: $\boldsymbol{a}_1 = k_1\boldsymbol{m}_1 + k_2\boldsymbol{m}_2 + \ldots + k_s\boldsymbol{m}_s$. Отсюда уже следует, что в системе M должен быть *хотя бы один вектор*, т. е. что первое утверждение теоремы в данном случае справедливо. Кроме того, так как $\boldsymbol{a}_1 \neq 0$, хотя бы один из коэффициентов $k_1, k_2, \ldots, k_s$, например k_1, должен быть отличным от нуля (в противном случае линейная комбинация равнялась бы нулю, а не вектору $\boldsymbol{a}_1$). Если $k_1 \neq 0$, то записанное выше равенство можно переписать так: $\boldsymbol{m}_1 = \frac{1}{k_1}\boldsymbol{a}_1 - \frac{k_2}{k_1}\boldsymbol{m}_2 - \ldots - \frac{k_s}{k_1}\boldsymbol{m}_s$.

Но теперь видно непосредственно, что система M', полученная заменой в системе M вектора $\boldsymbol{m}_1$ вектором $\boldsymbol{a}_1$, будет эквивалентна системе M.

В самом деле, если мы возьмём любой вектор системы M, то могут представиться два и только два случая:

а) Выбранный вектор $\boldsymbol{x}$ отличен от $\boldsymbol{m}_1$; но тогда он принадлежит также и системе M', так как мы его не выбрасывали при замене. Равенство $\boldsymbol{x} = 1\boldsymbol{x}$ показывает, таким образом, что вектор $\boldsymbol{x}$ является линейной комбинацией векторов системы M'.

б) Выбранный вектор $\boldsymbol{x} = \boldsymbol{m}_1$. Но в этом случае имеющееся в нашем распоряжении равенство $\boldsymbol{m}_1 = \frac{1}{k_1}\boldsymbol{a}_1 - \frac{k_2}{k_1}\boldsymbol{m}_2 - \ldots - \frac{k_s}{k_1}\boldsymbol{m}_s$ также показывает, что он является линейной комбинацией векторов системы M'.

Наоборот, если взять любой вектор $\boldsymbol{x}$ системы M', то опять возможны два случая: а) $\boldsymbol{x} \neq \boldsymbol{a}_1$ и б) $\boldsymbol{x} = \boldsymbol{a}_1$. В этих случаях выражение вектора $\boldsymbol{x}$ в виде линейной комбинации векторов системы M даётся, соответственно, равенствами $\boldsymbol{x} = \boldsymbol{x}$ и $\boldsymbol{x} = \boldsymbol{a}_1 = k_1\boldsymbol{m}_1 + \ldots + k_s\boldsymbol{m}_s$. Доказательство эквивалентности систем M и M' закончено.

Переходя к случаю произвольного $n > 1$, предположим, что наша теорема уже доказана в случае, когда данная линейно независимая

система содержит $n-1$ векторов. В этом предположении докажем справедливость обоих утверждений для системы n векторов.

В самом деле, если исключить из данной системы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, \boldsymbol{a}_n$ последний вектор $\boldsymbol{a}_n$, то останется система из $n-1$ векторов, которая попрежнему будет линейно независимой и такой, что каждый её вектор будет линейной комбинацией векторов системы M. Отсюда в силу предположения индукции следует, что число векторов системы $M \geqslant n-1$ и что $n-1$ векторов этой системы можно заменить векторами $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$ так, что полученная после такой замены система M'' будет эквивалентна системе M. Рассмотрим теперь вектор $\boldsymbol{a}_n$. Он является линейной комбинацией векторов системы M по условию теоремы, так как системы M и M'' эквивалентны, он будет также линейной комбинацией векторов последней системы. В число этих векторов могут войти и векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$, так что $\boldsymbol{a}_n = k_1\boldsymbol{a}_1 + \ldots + k_{n-1}\boldsymbol{a}_{n-1} + l_1\boldsymbol{m}_1 + \ldots + l_s\boldsymbol{m}_s$, где $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_s$ — могущие присутствовать в этом выражении векторы первоначальной системы M. Покажем, что хотя бы один такой вектор должен входить в выражение $\boldsymbol{a}_n$ с коэффициентом, отличным от нуля.

В самом деле, если все коэффициенты $l_1, \ldots, l_s$ равны нулю (или векторов $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_s$ вообще нет), то вектор $\boldsymbol{a}_n$ оказывается линейной комбинацией векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$, вопреки линейной независимости системы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}, \boldsymbol{a}_n$.

Доказанный факт обнаруживает, что в системе M'' найдётся хотя бы один вектор $\boldsymbol{m}$, отличный от векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}$. Но это означает, что число векторов в системе M было $\geqslant n$. Кроме того, повторение доказательства для случая $n=1$ даёт, что один из векторов $\boldsymbol{m}_i$ системы M'' может быть заменён вектором $\boldsymbol{a}_n$ так, чтобы полученная после такой замены система M' была эквивалентна системе M''. Учитывая теперь, что системы M и M'' эквивалентны и что система M' в конечном счёте получена из системы M заменой n её векторов векторами $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, убеждаемся в справедливости и второго утверждения теоремы.

Подчеркнём, что в теореме не утверждается возможность замены любых векторов системы M векторами $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Из доказательства видно, что при каждом этапе замены можно заменять только тот вектор, который входит в рассматриваемое соотношение с отличным от нуля коэффициентом. Что касается того, какие именно векторы будут обладать этим свойством, то это в общем случае заранее предсказать нельзя. Проводимые в доказательстве рассуждения обнаруживали только, что хотя бы один такой вектор обязательно существует.

Содержание доказанной теоремы раскрывается теми следствиями, которые могут быть из неё получены. Прежде всего укажем такое:

Если каждый вектор системы M является линейной комбинацией конечного множества векторов $\boldsymbol{b}_1, \boldsymbol{b}_2, , \ldots, \boldsymbol{b}_m$, то

никакая линейно независимая часть системы M не может содержать больше m векторов.

В самом деле, если $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ — линейно независимая часть системы M, то к ней и к системе векторов $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ может быть применена теорема о замене, из которой следует, что $n \leqslant m$.

Этот результат может быть применён, в частности, к пространству числовых векторов: в § 2 мы видели, что каждый n-мерный числовой вектор может быть представлен линейной комбинацией фиксированных векторов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Следовательно, в силу только что отмеченного следствия теоремы о замене *никакая линейно независимая система n-мерных числовых векторов не может содержать более n векторов.*

Другим важным следствием теоремы о замене является следующее предложение:

Если две конечные системы векторов $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ и $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p$ линейно независимы и эквивалентны, то число элементов в обеих системах одно и то же: $m = p$.

В самом деле, эквивалентность систем и линейная независимость каждой из них позволяют использовать теорему о замене дважды: один раз первая из систем принимается за множество $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, а другая за M, а другой раз — наоборот. Это даёт два неравенства $m \leqslant p$ и $p \leqslant m$ для чисел элементов в системах. Доказываемое равенство из них следует непосредственно.

Рассмотрим теперь *какое-либо* множество векторов данного нам пространства и будем выбирать из него всевозможными способами конечные линейно независимые части. При этом логически возможны два случая: либо можно выбрать линейно независимые части, содержащие сколь угодно большое число векторов, либо же число векторов в каждой из таких частей никогда не будет превосходить некоторого числа n. Сделанное только что замечание об n-мерном числовом пространстве показывает, что в случае числовых векторов будет иметь место как раз второй случай.

То, что и первый случай не является только *логической возможностью*, можно увидеть, например, в случае пространства всех многочленов F_∞ (см. пример 4 § 8). Степени x: $x, x^2, \ldots, x^n, \ldots$, являются «векторами» этого пространства. Их линейными комбинациями являются многочлены $k_1x + k_2x^2 + \ldots + k_nx^n$, причём коэффициенты линейной комбинации являются просто коэффициентами этих многочленов. Отсюда видно, что нулевой многочлен (нулевой вектор нашего пространства!) мы можем получить, только взяв все коэффициенты равными нулю. Следовательно, «векторы» $x, \ldots, x^n, \ldots$ — линейно независимы, и мы можем из них выбрать конечную линейно независимую систему, содержащую сколь угодно большое число элементов.

Мы примем такое определение:

Рангом множества M векторов называется максимальное число векторов в линейно независимых частях этого множества.

Если ранга в смысле этого определения не существует (как в только что рассмотренном пространстве F_∞), то мы будем говорить, что ранг бесконечен.

Пусть M — какое-либо множество векторов пространства L (может быть, M совпадает со всем пространством) и пусть $a_1, \ldots, a_n$ — какая-либо конечная линейно независимая система векторов множества M. Будем называть её *максимальной линейно независимой системой во множестве M*, если добавление к ней любого вектора нашего множества л и ш а е т её свойства быть линейно независимой.

Для таких систем имеет место следующая

Т е о р е м а. *Линейно независимая система $a_1, a_2, \ldots, a_n$ векторов тогда и только тогда является максимальной линейно независимой системой во множестве M, когда она эквивалентна всему множеству. Если $a_1, \ldots, a_n$ и $b_1, \ldots, b_m$ — две максимальные линейно независимые системы векторов множества M, то они содержат одно и то же число элементов: $m = n$.*

Легко видеть, что второе утверждение непосредственно вытекает из первого в силу доказанного выше следствия теоремы о замене, относящегося к эквивалентным линейно независимым системам. Доказательство же первого утверждения можно провести так:

1) Пусть система $a_1, \ldots, a_n$ является м а к с и м а л ь н о й линейно независимой системой векторов множества M; все её векторы линейно выражаются через векторы множества M. С другой стороны, если x — любой вектор из M, то система векторов $a_1, \ldots, a_n, x$ будет уже линейно зависимой. Поэтому должна существовать равная нулю линейная комбинация этих векторов

$$k_1 a_1 + \ldots + k_n a_n + kx = 0,$$

имеющая хотя бы один отличный от нуля коэффициент. Но легко видеть, что k не может равняться нулю, так как в этом случае был бы отличным от нуля один из других коэффициентов и векторы $a_1, \ldots, a_n$ были бы линейно зависимыми, вопреки предположению. Это же в свою очередь означает, что вектор

$$x = -\frac{k_1}{k} a_1 - \ldots - \frac{k_n}{k} a_n.$$

Так как x — произвольный вектор из M, то векторы из M линейно выражаются через векторы системы $a_1, \ldots, a_n$.

2) Наоборот, пусть линейно независимая система векторов $a_1, \ldots, a_n$, множества M эквивалентна всему множеству. Тогда любой вектор x множества M будет некоторой линейной комбинацией векторов данной системы:

$$x = k_1 a_1 + \ldots + k_n a_n \quad \text{или} \quad 1x - k_1 a_1 - \ldots - k_n a_n = 0$$

Так как эта линейная комбинация равна нулю, а не все её коэффициенты равны нулю, то система $a_1, \ldots, a_n, x$ не может быть линейно независимой, что и требовалось доказать.

Понятие ранга множества векторов приобретает особенно наглядный смысл в случае множеств векторов на плоскости или в обычном трёхмерном пространстве. Как легко видеть, в этих случаях ранг равен, соответственно, 2 и 3, т. е. совпадает с тем числом, которое обычно называют *числом измерений*. Рассматривая совокупность векторов, лежащих на одной прямой линии, мы без труда увидим, что её ранг равен 1, т. е. опять-таки совпадает с числом измерений, которое обычно приписывается прямой.

Разобранный пример делает естественным следующее определение размерности произвольного векторного пространства:

Размерностью векторного пространства называется число элементов в максимальной линейно независимой системе векторов этого пространства.

Доказанная выше теорема позволяет утверждать, что это число не зависит от выбора максимальной линейно независимой системы векторов.

Замечая, что векторы $e_1, \ldots, e_n$ числового n-мерного пространства образуют, очевидно, максимальную линейно независимую систему, можем утверждать теперь, что размерность этого пространства в только что определённом смысле равна n. Этим оправдывается применявшееся нами раньше название «n-мерное» пространство.

§ 11. Подпространства

Если мы рассмотрим векторы обычного трёхмерного пространства, лежащие на какой-либо плоскости, то немедленно обнаружим, что суммы этих векторов и произведения их на действительные числа являются снова векторами той же плоскости. Это обстоятельство наводит на мысль придать нашей терминологии, относящейся к общему случаю любых векторных пространств, ещё бóльшую геометричность, введя такое определение:

Подпространством данного векторного пространства называется любое множество L векторов этого пространства, обладающее двумя следующими свойствами:

1. *Если векторы a и b принадлежат этому множеству, то и их сумма также принадлежит L.*

2. *Если вектор a принадлежит множеству L, то и его произведение ka на любое число из поля K также принадлежит L.*

Нетрудно видеть, что каждое подпространство само является векторным пространством в смысле общего определения, введённого в § 8: для двух его элементов определена их сумма, определено произведение любого из его элементов на число из поля K и свойства этих операций, выраженные аксиомами I—V,

также выполнены, так как они выполняются во всём пространстве. Таким образом, *всё, что мы выше говорили о векторных пространствах, автоматически относится и к их подпространствам.*

Однако при рассмотрении подпространств обнаруживается целый ряд новых явлений, связанных с их, так сказать, «взаимным расположением».

Пусть L_1 и L_2 — два подпространства одного и того же векторного пространства L. *Пересечением* этих подпространств естественно назвать *совокупность векторов пространства, принадлежащих одновременно обоим подпространствам.* Это наименование находится в согласии как с наглядными геометрическими представлениями, так и с общим определением теории множеств, в которой, как известно, пересечением любых множеств называют совокупность их общих элементов. Если мы рассмотрим совокупности векторов обычного трёхмерного пространства, лежащих на двух плоскостях, как подпространства, то пересечением их в нашем смысле будет совокупность векторов, лежащих на прямой, по которой пересекаются эти плоскости. Эта совокупность оказывается сама подпространством. Если аналогичным образом рассмотреть совокупности векторов, лежащих на пересекающихся прямой и плоскости, то пересечение этих подпространств будет состоять только из одного нулевого вектора: только нулевой вектор можно рассматривать как лежащий одновременно на нашей плоскости и на прямой. Однако нулевой вектор *сам по себе образует подпространство*: ведь сложение нулевого вектора с самим собой и умножение его на любое число дают снова нулевой вектор. Эти совершенно наглядные соображения приводят к предположению о справедливости следующей общей теоремы.

Теорема. *Пересечение двух подпространств любого пространства само является подпространством.*

Действительно, пусть L_1 и L_2 — данные подпространства пространства L. Если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ содержатся в пересечении этих подпространств, то они будут содержаться также и в каждом из них в отдельности, например в L_1. Но так как L_1 есть подпространство, то сумма этих векторов $\boldsymbol{a}+\boldsymbol{b}$ и произведение $k\boldsymbol{a}$ одного из них на любое число также принадлежит подпространству. Из тех же соображений следует, что эти сумма и произведение принадлежат также и другому подпространству L_2, а значит, и пересечению этих подпространств. Наше утверждение доказано.

Конечно, можно говорить также и о пересечении любого числа подпространств. Оно будет также подпространством рассматриваемого пространства.

Подпространства могут содержаться одно в другом: *это выражение будет означать просто, что каждый вектор, принадлежащий первому из них, является также и вектором второго.* Например, пересечение двух или большего числа подпространств

содержится в каждом из первоначально заданных подпространств. Ясно, что если подпространство L_1 пространства L содержится в другом подпространстве L_2 того же пространства, то L_1 может рассматриваться так же, как подпространство L_2. Отметим также, что любое пространство может рассматриваться как подпространство в себе самом: это вполне согласуется с принятым нами определением. Если нам нужно подчеркнуть, что рассматриваемое подпространство не совпадает со всем пространством, то мы будем называть его *собственным подпространством.*

Легко видеть, что если задано произвольное множество M векторов пространства L, то совокупность всех их линейных комбинаций образует уже подпространство в L: ведь сумма двух линейных комбинаций векторов множества M и произведение любой из этих линейных комбинаций на любое число рассматриваемого поля K будут снова линейными комбинациями векторов множества M. Это подпространство называется *подпространством, порождаемым данным множеством векторов.*

Наибольший интерес для нас будут представлять случаи, когда рассматриваемое множество векторов конечно.

В случае векторного пространства элементарной геометрии подпространство, порождаемое одним вектором, является совокупностью векторов, лежащих на определяемой этим вектором прямой линии. То же самое подпространство порождают и два вектора этой прямой, если хотя бы один из них отличен от нуля. Однако если мы возьмём два вектора, не лежащих на одной прямой, то порождаемое ими подпространство будет уже плоскостью. Наконец, если мы возьмём три вектора, не лежащих на одной плоскости, то порождаемое ими подпространство будет попросту совпадать со всем пространством.

Можно указать ряд подпространств в тех пространствах, которые были определены в примерах 1—4 § 8. Например, в пространстве F_∞, уже рассматривавшемся в предыдущем параграфе, в качестве подпространства содержится *совокупность многочленов, не содержащих нечётных степеней x.* Легко видеть, что размерность этого подпространства так же бесконечна, как и размерность самого пространства: ведь $1, x^2, \ldots, x^{2n}, \ldots$ линейно независимы (см. § 10). Однако в том же пространстве имеются и подпространства конечной размерности. Так, пространство F_n многочленов, степень которых не превышает n, содержится в F_∞ в качестве подпространства. То, что размерность этого подпространства конечна, ясно из того, что все многочлены степени n или меньше представляются линейными комбинациями «одночленов» $1, x, x^2, \ldots, x^n$, также являющихся «векторами» нашего пространства.

Доказанные в предыдущем параграфе теоремы дают возможность во многих случаях просто определить, какова размерность того или

иного пространства, как это сделано выше в случае пространства n-мерных числовых векторов.

Например, размерность пространства F_n многочленов степени $\leqslant n$ равна $n+1$, так как «векторы» этого пространства $1, x, \ldots, x^n$ линейно независимы и любой вектор (т. е. любой многочлен указанной степени) представляется их линейной комбинацией.

Между размерностями пространства и его подпространства имеет место такое соотношение:

Размерность любого подпространства не превосходит размерности пространства. Если размерность пространства конечна, то размерность любого собственного подпространства строго меньше размерности пространства.

В самом деле, первое утверждение очевидно, так как всякая линейно независимая система векторов подпространства будет также линейно независимой системой векторов всего пространства. Для доказательства второго утверждения предположим противное. Возьмём некоторую максимальную линейно независимую систему векторов подпространства. Число входящих в неё векторов равно в точности размерности подпространства, а следовательно, и размерности пространства, так как эти размерности предположены равными. По этой причине она будет также максимальной линейно независимой системой векторов всего пространства. Но в таком случае эта система, по доказанной в предыдущем параграфе теореме должна быть эквивалентна всему пространству. Это, в частности, означает, что любой вектор пространства является линейной комбинацией векторов нашей системы, которые все принадлежат подпространству. Принимая теперь во внимание, что любая линейная комбинация векторов подпространства принадлежит к этому же подпространству, получаем, что любой вектор пространства принадлежит рассматриваемому подпространству, так что подпространство не будет собственным.

§ 12. Применение к системам уравнений

Мы сделаем сейчас небольшое отступление от изложения общей теории, чтобы дать понять, каким образом развиваемые здесь соображения могут оказаться полезными для исследования систем уравнений.

Пусть дана некоторая система линейных уравнений

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m &= b_n. \end{aligned}\right\} \qquad (1)$$

Здесь мы не предполагаем, что число уравнений системы равно числу неизвестных: оно может быть как больше, так и меньше этого числа.

Требуется найти способ узнать, имеет ли данная система решения и (если она их имеет) каково число решений. Наконец, желательно также указать способ нахождения всех решений системы.

Все эти вопросы, как мы увидим, легко сводятся к некоторым вопросам о векторах, тесно примыкающим к только что рассмотренным.

Действительно, будем рассматривать каждый столбец коэффициентов при одном и том же неизвестном, а также и столбец свободных членов как n-мерные числовые векторы. Обозначим эти векторы, соответственно, через $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ и $\boldsymbol{b}$. Тогда система (1) может быть записана в виде одного уравнения

$$\boldsymbol{a}_1 x_1 + \boldsymbol{a}_2 x_2 + \ldots + \boldsymbol{a}_m x_m = \boldsymbol{b}. \qquad (2)$$

Любое решение системы (1), т. е. любая совокупность значений неизвестных, удовлетворяющая системе (1), будет также удовлетворять уравнению (2), и наоборот. Это следует просто из того, что уравнение (2) является лишь иной записью системы (1).

Это может быть сформулировано другими словами так:

Система уравнений (1) *имеет решение тогда и только тогда, когда вектор* $\boldsymbol{b}$ *является линейной комбинацией векторов* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$.

Как мы увидим дальше, особенно важной для нас будет следующая формулировка того же по существу результата:

Система уравнений (1) *тогда и только тогда имеет решение, когда ранги систем векторов* $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ и $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m, \boldsymbol{b}$ *равны.*

Действительно, ранг системы векторов есть число векторов максимальной линейно независимой части этой системы. Если ранги систем $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ и $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m, \boldsymbol{b}$ равны, то максимальная линейно независимая часть системы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ будет также максимальной линейно независимой частью системы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m, \boldsymbol{b}$. Но по доказанной в предыдущем параграфе теореме максимальная линейно независимая часть любого множества эквивалентна этому множеству векторов. Так как в нашем случае она будет эквивалентна обеим рассматриваемым системам, то эти последние будут эквивалентны между собой, а следовательно, вектор $\boldsymbol{b}$ будет линейной комбинацией векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$, и система имеет решение.

Наоборот, если решение системы (1) существует, то вектор $\boldsymbol{b}$ оказывается линейной комбинацией векторов системы $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$. Отсюда сразу видно, что каждый из векторов любой из рассматри-

ваемых систем будет линейной комбинацией векторов другой системы (сомнения могли бы возникнуть лишь в отношении вектора b), т. е. эти системы эквивалентны.

Полученное условие существования решения системы (1) может быть формулировано и в терминах подпространств. При этом резче бросится в глаза совпадение полученного результата с тем, который был формулирован в § 1 *для случая системы двух уравнений с двумя неизвестными.* Разница состоит лишь в том, что здесь мы одной формулировкой охватываем все возможные случаи. Эта формулировка такова:

Система (1) *тогда и только тогда имеет решение, когда вектор* b *содержится в подпространстве, порождаемом векторами* $a_1, a_2, \ldots, a_m$.

Читателю предлагается самому установить справедливость этой формулировки и привести её в связь с формулировкой, данной в § 1.

Исходя из изложенных соображений, можно получить также условие единственности решения системы (1).

Решение системы (1) *будет единственным в том и только в том случае, когда векторы* $a_1, \ldots, a_m$ *линейно независимы* (мы предполагаем здесь, что решение существует).

В самом деле, пусть векторы $a_1, \ldots, a_m$ линейно независимы. Если бы система (1), значит и уравнение (2), имела бы два различных решения $x_1, \ldots, x_m$ и $x_1', \ldots, x_m'$, то из равенств

$$a_1x_1 + \ldots + a_mx_m = b,$$

$$a_1x_1' + \ldots + a_mx_m' = b$$

получалось бы равенство

$$a_1(x_1 - x_1') + \ldots + a_m(x_m - x_m') = 0.$$

Но линейная комбинация линейно независимых векторов может быть равна нулю только в том случае, когда её коэффициенты $(x_1 - x_1')$, $(x_2 - x_2')$, $\ldots$, $(x_m - x_m')$ равны нулю. Это противоречит тому, что взятые нами решения различны.

Наоборот, если векторы $a_1, \ldots, a_m$ линейно зависимы, то найдётся их линейная комбинация, не все коэффициенты которой равны нулю, равная нулю: $a_1k_1 + a_2k_2 + \ldots + a_mk_m = 0$. Тогда из любого решения $x_1, \ldots, x_m$ уравнения (2) можно образовать новое решение $x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m$, отличное от первоначального. В самом деле,

$$a_1(x_1 + k_1) + \ldots + a_m(x_m + k_m) =$$

$$= a_1x_1 + \ldots + a_mx_m + a_1k_1 + \ldots + a_mk_m = a_1x_1 + \ldots + a_mx_m = b,$$

т. е. вновь образованная система значений неизвестных действительно удовлетворяет уравнению (2).

Мы ограничимся пока этими результатами. Их мы будем в состоянии ещё уточнить, после того как найдём способ вычисления ранга систем векторов.

§ 13. Базис пространства. Координаты

В конце § 11 мы видели, что любая максимальная линейно независимая система векторов какого-либо множества эквивалентна всему этому множеству. В частности, максимальная линейно независимая система $e_1, \ldots, e_n$ векторов пространства L эквивалентна всему пространству. Это означает, что любой вектор x пространства может быть представлен в виде линейной комбинации векторов нашей системы:

$$x = e_1x_1 + \ldots + e_nx_n, \tag{1}$$

где x_i — некоторые числа рассматриваемого числового поля.

Но равенство (1) можно рассматривать как уравнение относительно неизвестных коэффициентов $x_1, \ldots, x_n$. Полученное в предыдущем параграфе условие единственности решения такого уравнения в рассматриваемом случае выполнено, так как векторы $e_1, \ldots, e_n$ линейно независимы.

Подобное положение, замеченное нами в § 1 в случае векторов плоскости, дало возможность ввести для векторов плоскости координаты. Это же может быть теперь сделано в случае произвольного векторного пространства L над числовым полем K.

Будем называть *любую максимальную линейно независимую систему векторов* $e_1, \ldots, e_n$ пространства L его *базисом* (или *базой*).

Если задан такой базис, то равенство (1) сопоставляет с каждым вектором систему чисел $x_1, \ldots, x_n$. Эти числа мы будем называть *координатами вектора* x *относительно базиса* $e_1, \ldots, e_n$. Ясно, что любые значения координат будут соответствовать некоторому вектору пространства L (конечно, если они берутся из того числового поля, на котором определено наше пространство).

Введение координат, как и в обычном случае, даёт возможность сводить исследование систем векторов к исследованию систем чисел. Однако, в случае обычной аналитической геометрии, мы получаем таким образом только аппарат, помогающий часто более просто решать задачи, которые могут быть решены и другим геометрическим способом. В общем же случае такое сведéние к числам будет единственным способом, которым мы вообще в состоянии решать задачи, так как никаких «геометрических» методов в нашем распоряжении нет.

Следует отметить, что термином координаты мы уже пользовались выше для числовых векторов: координатами вектора $a = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$ были названы сами числа $a_1, \ldots, a_n$, составляющие вектор. Вспоминая, что в этом случае вектор может быть представлен в виде $a = a_1e_1 + \ldots + a_ne_n$ и что «единичные» векторы $e_1, \ldots, e_n$ линейно независимы, можно сказать, что теперь было бы правильнее назвать числа $a_1, \ldots, a_n$ координатами вектора a *по отношению к базису* $e_1, \ldots, e_n$ n-мерного числового пространства. Такая оговорка действительно необходима, так как мы можем рассматривать вместо базиса $e_1, \ldots, e_n$ любой другой базис $e'_1, \ldots, e'_n$ (в нём, конечно, содержится то же самое число векторов), причём координаты того же вектора по отношению к новому базису будут иметь уже другие значения.

То обстоятельство, что координаты вектора не являются чем-то абсолютным, а зависят от выбора базиса, тотчас же ставит перед нами такие вопросы:

1) Как узнать, можно ли данную систему векторов принять за базис пространства?

2) Как связаны между собою координаты одного и того же вектора относительно двух разных базисов пространства?

Рассмотрим, прежде всего, первый вопрос. Пусть в пространстве L задан какой-либо базис $e_1, \ldots, e_n$ и пусть дана некоторая система векторов $e'_1, e'_2, \ldots, e'_m$. Каждый из векторов этой системы может быть однозначно представлен линейной комбинацией векторов заданного базиса. Коэффициенты полученных линейных комбинаций мы будем обозначать одной и той же буквой c_{ij} с двумя индексами, второй из которых соответствует номеру вектора e'_j, а первый — тому из векторов e_i, при котором коэффициент стоит. Таким образом,

$$\left.\begin{aligned} e'_1 &= e_1c_{11} + e_2c_{21} + \ldots + e_nc_{n1}, \\ e'_2 &= e_1c_{12} + e_2c_{22} + \ldots + e_nc_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ e'_m &= e_1c_{1m} + e_2c_{2m} + \ldots + e_nc_{nm}. \end{aligned}\right\} \qquad (2)$$

Коэффициенты c_{ij}, естественно, располагаются в прямоугольную таблицу, или матрицу

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{pmatrix}, \qquad (3)$$

которую будем называть *матрицей перехода* от базиса $e_1, \ldots, e_n$ к системе $e'_1, \ldots, e'_m$.

Иными словами, матрицей перехода от базиса $e_1, e_2, \ldots, e_n$ к системе $e'_1, e'_2, \ldots, e'_m$ мы назвали матрицу, столбцы которой составлены из координат векторов системы $e'_1, \ldots, e'_n$ относительно данного базиса. Эта матрица однозначно определена при заданной системе $e'_1, \ldots, e'_m$ и сама определяет эту систему (конечно, при заданном базисе $e_1, \ldots, e_n$). Стоящий перед нами вопрос может быть теперь формулирован точнее следующим образом: *как по матрице перехода* (3) *узнать, является ли система векторов* $e'_1, \ldots, e'_m$ *базисом пространства* L?

Ответ на этот вопрос даётся такой теоремой:

Теорема. *Для того чтобы система* $e'_1, e'_2, \ldots, e'_m$ *была базисом пространства* L, *необходимо и достаточно, чтобы матрица перехода* (3) *была квадратной и чтобы составленный из неё определитель был отличен от нуля.*

В самом деле, система $e'_1, \ldots, e'_m$ только тогда может быть базисом пространства L, когда число векторов этой системы равно n (ведь число элементов векторов в любых базисах пространства L одно и то же). Для доказательства того, что определитель матрицы (3) отличен от нуля, если $m = n$ и система $e'_1, \ldots, e'_m$ является базисом, воспользуемся соображениями § 2—4.

Для любых векторов

$$\left.\begin{aligned} x_1 &= e_1x_{11} + e_2x_{21} + \ldots + e_nx_{n1}, \\ x_2 &= e_1x_{12} + e_2x_{22} + \ldots + e_nx_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= e_1x_{1n} + e_2x_{2n} + \ldots + e_nx_{nn}, \end{aligned}\right\} \tag{4}$$

пространства L положим:

$$F(x_1, x_2, \ldots, x_n) = \begin{vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \ldots & x_{nn} \end{vmatrix}. \tag{5}$$

Этим определена некоторая функция от n векторов пространства, не равная тождественно нулю, так как её значение $F(e_1, \ldots, e_n) = 1$ по самому определению этой функции. С другой стороны, если положить $x_1 = e'_1, \ldots, x_n = e'_n$, то значение функции F обратится в интересующий нас определитель матрицы (3). Обратим теперь внимание на то, что наша функция обладает свойствами А) и Б) определения, сформулированного в § 2. Поэтому для неё будут верны также и следствия этих свойств, доказанные в § 2, в частности *распределительный закон и свойство менять знак при перестановке аргументов*. Если система $e'_1, \ldots, e'_n$ является базисом пространства, то эти свойства позволяют получить другое выраже-

ние той же функции: векторы $x_1, \ldots, x_n$ запишутся в выражении через базис $e'_1, \ldots, e'_n$ таким образом:

$$\left.\begin{aligned} x_1 &= e'_1x'_{11} + e'_2x'_{21} + \ldots + e'_nx'_{n1}, \\ x_2 &= e'_1x'_{12} + e'_2x'_{22} + \ldots + e'_nx'_{n2}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n &= e'_1x'_{1n} + e'_2x'_{2n} + \ldots + e'_nx'_{nn}. \end{aligned}\right\} \qquad (6)$$

Подставляя эти выражения в $F(x_1, x_2, \ldots, x_n)$ и пользуясь распределительным законом и правилом перестановки аргументов, получим, как и в § 4, выражение

$$F(x_1, x_2, \ldots, x_n) = \begin{vmatrix} x'_{11} & x'_{12} & \ldots & x'_{1n} \\ x'_{21} & x'_{22} & \ldots & x'_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x'_{n1} & x'_{n2} & \ldots & x'_{nn} \end{vmatrix} F(e'_1, e'_2, \ldots, e'_n). \qquad (7)$$

Нужный нам результат следует из самой возможности такого выражения: если бы значение нашего определителя, т. е. значение $F(e'_1, \ldots, e'_n)$, было равно нулю, то выражение (7) давало бы тождественный нуль для любых векторов $x_1, \ldots, x_n$, вопреки определению функции $F(x_1, \ldots, x_n)$.

Таким образом, оба условия теоремы являются необходимыми для того, чтобы система векторов $e'_1, \ldots, e'_n$ была базисом пространства L.

Предположим теперь, что эти условия выполнены: число векторов e'_j равно n и указанный определитель не равен нулю. Система $e'_1, \ldots, e'_n$ могла бы не быть базисом пространства L только в том случае, когда один из её векторов был бы линейной комбинацией остальных, например

$$e'_1 = \sum_{i=2}^{n} k_i e'_i.$$

Но в этом случае получаем равенство

$$F(e'_1, e'_2, \ldots, e'_n) = F\left(\sum_{i=2}^{n} k_i e'_i, e'_2, \ldots, e'_n\right) =$$

$$= \sum_{i=2}^{n} k_i F(e'_i, e'_2, \ldots, e'_n) = 0,$$

противоречащее сделанному предположению. Теорема доказана.

Второй из поставленных выше вопросов также легко решается: пусть некоторый вектор x имеет относительно базиса $e'_1, \ldots, e'_n$ координаты $x'_1, x'_2, \ldots, x'_n$. Тогда $x = e'_1x'_1 + e'_2x'_2 + \ldots + e'_nx'_n$.

Подставляя сюда выражения (2) для векторов нового базиса через старый, получим равенство

$$x=(e_1c_{11}+\ldots+e_nc_{n1})x'_1+\ldots+(e_1c_{1n}+\ldots+e_nc_{nn})x'_n=$$
$$=e_1(c_{11}x'_1+c_{12}x'_2+\ldots+c_{1n}x'_n)+\ldots+e_n(c_{n1}x'_1+\ldots+c_{nn}x'_n).$$

Сравнивая это выражение с первоначальным выражением (1) вектора x через векторы базиса $e_1, \ldots, e_n$ и учитывая, что координаты вектора определены (при заданном базисе) однозначно, получим систему равенств

$$\left.\begin{aligned} x_1&=c_{11}x'_1+c_{12}x'_2+\ldots+c_{1n}x'_n,\\ x_2&=c_{21}x'_1+c_{22}x'_2+\ldots+c_{2n}x'_n,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ x_n&=c_{n1}x'_1+c_{n2}x'_2+\ldots+c_{nn}x'_n, \end{aligned}\right\} \qquad (8)$$

выражающих «старые» координаты $x_1, \ldots, x_n$ вектора x через его «новые» координаты $x'_1, \ldots, x'_n$.

Равенства (8) при заданных «старых» координатах $x_1, \ldots, x_n$ можно рассматривать как систему n уравнений с n неизвестными. Определитель этой системы есть определитель матрицы (3) и поэтому отличен от нуля, так что система (8) допускает единственное решение (в силу доказанной в главе I основной теоремы о системах линейных уравнений). Таким образом, *знание координат вектора относительно любого из рассматриваемых базисов достаточно для определения их относительно другого базиса, если известна матрица перехода, связывающая базисы.*

§ 14. Ранг произвольной системы векторов

Полученный в предыдущем параграфе результат даёт возможность узнавать, равен ли ранг системы векторов размерности рассматриваемого пространства. Этот результат может быть теперь обобщён так, что в нашем распоряжении окажется способ, позволяющий *вычислить ранг произвольной системы векторов.*

Рассмотрение любых систем векторов сводится к рассмотрению базиса пространства с помощью простой теоремы, непосредственно получающейся из доказанной выше теоремы о замене.

Теорема. *Если $e_1, \ldots, e_n$ есть базис n-мерного пространства L, а $e'_1, e'_2, \ldots, e'_m$ — произвольная линейно независимая система векторов того же пространства, то эта система может быть дополнена некоторыми из векторов $e_1, \ldots, e_n$ до базиса пространства L.*

Действительно, векторы $e'_1, \ldots, e'_m$ являются линейными комбинациями векторов базиса $e_1, \ldots, e_n$. Поэтому в силу того, что векторы $e'_1, \ldots, e'_m$ линейно независимы, и в силу теоремы о замене часть векторов базиса $e_1, \ldots, e_n$ (если $m=n$, то все

векторы базиса) может быть заменена векторами $e'_1, \ldots, e'_m$ так, что вновь полученная система векторов будет эквивалентна первоначальному базису. Это значит, что любой вектор пространства может быть записан в виде линейной комбинации векторов вновь полученной системы. Если бы эта система не была линейно независимой, то из неё можно было бы удалить хотя бы один вектор без потери только что указанного свойства. Однако векторы n-мерного пространства не могут быть все выражены линейными комбинациями меньшего числа векторов, чем n. Таким образом, векторы системы, полученной после замены, должны быть все линейно независимы, и эта система будет базисом пространства.

Как видно из проведённого построения, полученную систему можно рассматривать так же, как систему $e'_1, \ldots, e'_m$, к которой, быть может, добавлены некоторые из векторов заданного базиса $e_1, \ldots, e_n$.

Следует отметить, что если бы система $e'_1, \ldots, e'_m$ была *линейно зависимой*, то добавление к ней других векторов *не может дать базиса пространства*: любая часть базиса, как и любой линейно независимой системы векторов, должна быть сама линейно независимым множеством.

Доказанную теорему и сделанное замечание применим к произвольной системе векторов

$$\left.\begin{aligned} e'_1 &= e_1 c_{11} + \ldots + e_n c_{n1}, \\ e'_2 &= e_1 c_{12} + \ldots + e_n c_{n2}, \\ &\cdots\cdots\cdots\cdots \\ e'_m &= e_1 c_{1m} + \ldots + e_n c_{nm} \end{aligned}\right\} \quad (m \leqslant n). \tag{1}$$

Матрица перехода от базиса $e_1, \ldots, e_n$ к системе $e'_1, \ldots, e'_m$ запишется так:

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{pmatrix}. \tag{2}$$

Присоединим теперь к системе векторов $e'_1, \ldots, e'_m$ ещё $n - m$ произвольных различных векторов базиса $e_1, \ldots, e_n$. Полученная система n векторов будет соответствовать матрице перехода, образующейся из матрицы (1) приписыванием к ней ещё $n - m$ столбцов, в каждом из которых все элементы, кроме одного, равны нулю, а этот единственный элемент равен единице (разложение любого из векторов базиса имеет вид $e_i = e_1 \cdot 0 + \ldots + e_i \cdot 1 + \ldots + e_n \cdot 0$). Следует подчеркнуть, что единицы будут стоять в различных строчках матрицы.

Согласно результату предыдущего параграфа векторы расширенной системы будут образовывать базис пространства в том и только в том случае, если определитель новой матрицы перехода отличен от нуля.

Но если учесть то, что было сказано о строении матрицы перехода, мы сразу замечаем возможность разложить определитель этой матрицы по элементам последнего столбца (так как в последнем столбце имеется только одна единица, а остальные элементы равны нулю). Эту операцию можно продолжить и дальше, пока не дойдём до столбцов первоначальной матрицы (2). В результате мы получаем, что интересующий нас определитель совпадает, с точностью до знака, с определителем, полученным из матрицы (2) вычёркиванием тех строк, в которых стояли единицы приписываемых к этой матрице дополнительных столбцов.

Но это в связи с доказанной только что теоремой сразу даёт такой результат:

Теорема. *Векторы* (1) *линейно независимы тогда и только тогда, когда из матрицы* (2) *можно так вычеркнуть* $n - m$ *строк, чтобы определитель оставшейся квадратной матрицы был отличен от нуля.*

Действительно, если можно таким образом вычеркнуть строки, то, добавляя к векторам $e_1, \ldots, e_m$ векторы e_i, номера которых совпадают с номерами вычеркнутых строк, мы согласно доказанному получим систему векторов, для которой определитель соответствующей матрицы перехода отличен от нуля. Эта система будет в силу результата предыдущего параграфа базисом пространства, а следовательно, исходная система линейно независима.

Наоборот, если при вычёркивании любых $n - m$ строк матрицы (2) получаются квадратные матрицы, определители которых равны нулю, то, как бы мы ни добавляли к заданной системе векторы базиса $e_1, \ldots, e_m$, ни одна из получаемых систем n векторов не будет базисом пространства. Это возможно только в случае, когда векторы системы $e_1, \ldots, e_n$ линейно зависимы.

Нам остаётся только освободиться от сделанных ограничений, связанных с числом векторов системы $e_1, \ldots, e_m$. Результат, относящийся к любому числу этих векторов, получается из предыдущего без всякого труда. Однако для его формулировки полезно ввести один вспомогательный термин.

Пусть дана произвольная матрица

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{pmatrix}. \tag{3}$$

Она может не быть квадратной, причём число её столбцов может

быть как больше, так и меньше числа строк. Из матрицы (3) можно, вычёркивая некоторое число строк и некоторое число столбцов, различными способами образовать квадратные матрицы. Определители получаемых таким образом матриц называются *минорами матрицы* (3). Некоторые из этих миноров могут быть отличны от нуля, другие, наоборот, равны нулю.

Рангом матрицы (3) *мы будем называть наибольший порядок отличного от нуля минора этой матрицы.*

Ранг любой матрицы можно вычислить: для этого достаточно, например, вычислить все миноры матрицы и посмотреть, миноры какого порядка отличны от нуля. Вычисление может быть ещё упрощено, так как обращение всех миноров какого-либо порядка в нуль влечёт за собою обращение в нуль и всех миноров более высокого порядка (эти миноры, как и любые определители, могут быть выражены через миноры более низкого порядка!). Это делает ненужным вычисление всех миноров: достаточно обнаружить, что все миноры какого-либо порядка равны нулю, а миноры порядка на единицу меньшего не все равны нулю. Дальнейшие упрощения вычисления ранга будут указаны ниже.

Вычисление ранга произвольной системы векторов может быть теперь проведено, основываясь на следующей теореме:

Теорема. *Если* $e'_1, \ldots, e'_m$ *есть система любого числа векторов пространства* L, *то её ранг равен рангу матрицы перехода, связывающей эту систему с любым базисом нашего пространства.*

В самом деле, если ранг матрицы перехода равен r, то существует минор этой матрицы, отличный от нуля и имеющий порядок r. Пусть в этот минор входят элементы некоторых столбцов матрицы перехода. Взятые столбцы сами образуют матрицу перехода от базиса $e_1, \ldots, e_n$ к системе векторов $e'_{j_1}, \ldots, e'_{j_r}$, номера которых соответствуют номерам рассматриваемых столбцов. А так как из этих столбцов можно образовать отличный от нуля минор r-го порядка, то по доказанному выше векторы $e'_{j_1}, \ldots, e'_{j_r}$ линейно независимы.

Наоборот, если мы возьмём любую часть системы векторов $e'_1, \ldots, e'_m$, содержащую большее число элементов, чем r, то применение к этой части только что доказанного критерия линейной независимости даёт отрицательный результат, так что эта часть системы будет линейно зависимой.

Таким образом, максимальное число линейно независимых векторов, которое может быть выбрано из системы $e'_1, \ldots, e'_m$, точно равно рангу матрицы перехода.

Доказанная теорема позволяет получить одно любопытное следствие: до сих пор, говоря о числовых векторах, мы понимали под этими словами столбцы, составленные из чисел рассматриваемого

поля; но легко видеть, что в совокупности всех строк, содержащих по m чисел данного поля K, можно ввести такие же операции сложения и умножения на число, как и в случае столбцов. Тем самым множество строчек делается векторным пространством, и мы получаем право говорить о линейной независимости строк. В частности, можно говорить о линейной независимости строк некоторой матрицы. То обстоятельство, что миноры матрицы не изменяются от замены их строчек столбцами, позволяет на основании только что доказанной теоремы утверждать следующее:

Максимальное число линейно независимых столбцов, которое можно выбрать из данной матрицы, равно максимальному числу её линейно независимых строк.

§ 15. Решение произвольных систем линейных уравнений

Теоремы, доказанные в двух последних параграфах, позволяют не только придать окончательную форму результатам, полученным в § 12, но и получить приём, позволяющий фактически получать все решения любой системы линейных уравнений.

В самом деле, пусть дана система n уравнений с m неизвестными

$$\left.\begin{aligned} a_{11}x_1 + \ldots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + \ldots + a_{2m}x_m &= b_2, \\ \ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + \ldots + a_{nm}x_m &= b_n. \end{aligned}\right\} \qquad (1)$$

Столбцы коэффициентов при неизвестных и столбец свободных членов этих уравнений являются, как мы знаем, столбцами координат числовых векторов, обозначенных в § 12 через $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ и $\boldsymbol{b}$ (координаты относятся к базису из «единичных» векторов $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ числового пространства). Пользуясь тем, что необходимым и достаточным условием существования решения системы (1) является равенство рангов систем векторов $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ и $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m, \boldsymbol{b}$ (см. § 12), и связью между рангом системы векторов и рангом матрицы, можно иначе формулировать полученный в § 12 результат. Будем называть матрицу, составленную из коэффициентов при неизвестных, просто *матрицей системы*, а матрицу, получаемую из неё приписыванием к ней столбца свободных членов, — *расширенной* матрицей системы. Тогда необходимое и достаточное условие существования решения системы (1) примет такую форму:

Теорема[1]). *Для того чтобы система* (1) *имела хотя бы одно решение, необходимо и достаточно, чтобы ранг матрицы системы был равен рангу расширенной матрицы.*

---

[1]) Эта теорема называется иногда теоремой Кронекера-Капелли.

Если указанное условие выполнено, то следующий приём даёт возможность найти все решения системы (1). Пусть ранг матрицы системы и ранг расширенной матрицы равны r. Так как, в силу замечания в конце предыдущего параграфа, число r есть также максимальное число линейно независимых строк расширенной матрицы, то все её строки будут линейными комбинациями некоторых r строк. Но это означает, другими словами, что все уравнения системы (1) являются следствиями некоторых r уравнений системы (1) (все уравнения могут быть получены из уравнений, соответствующих линейно независимым строкам, умножением обеих частей каждого из последних уравнений на подходящие числовые множители и сложением соответствующих частей).

Таким образом, в рассматриваемом случае достаточно получить решение этих r уравнений, ибо любое их решение будет также решением остальных. Так как нумерация уравнений произвольна, можно предположить, что упомянутыми уравнениями являются первые уравнения системы

$$\left.\begin{aligned} a_{11}x_1 + \ldots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + \ldots + a_{2m}x_m &= b_2, \\ \ldots\ldots\ldots\ldots\ldots & \\ a_{r1}x_1 + \ldots + a_{rm}x_m &= b_r. \end{aligned}\right\} \tag{2}$$

Из существования решения всей данной системы вытекает, что решения такой «укороченной» системы подавно существуют. Поэтому ранг матрицы коэффициентов при неизвестных в новой системе будет также равен r. Последнее в свою очередь означает, что из столбцов этих коэффициентов можно выбрать r столбцов так, чтобы составленный из них определитель был отличен от нуля. Но нумерация столбцов зависит от нумерации неизвестных, которой мы можем распоряжаться по своему усмотрению. Следовательно, мы можем опять предположить, что такими столбцами являются первые r столбцов. Если число неизвестных равно r, то взятые столбцы исчерпывают все столбцы матрицы системы. В таком случае система (2) будет системой r уравнений с r неизвестными, определитель которой отличен от нуля. Но из § 7 мы знаем, что такого рода система всегда имеет единственное решение, определяемое формулами Крамера. Из проведённых рассуждений вытекает, что это решение и будет решением (также единственным) системы уравнений (1).

Если же r меньше числа неизвестных, то мы перенесём члены с неизвестными $x_{r+1}, \ldots, x_m$ в правую часть уравнений системы (2). Получим систему

$$\left.\begin{aligned} a_{11}x_1 + \ldots + a_{1r}x_r &= b_1 - a_{1,\,r+1}x_{r+1} - \ldots - a_{1m}x_m, \\ a_{21}x_1 + \ldots + a_{2r}x_r &= b_2 - a_{2,\,r+1}x_{r+1} - \ldots - a_{2m}x_m, \\ \ldots\ldots\ldots\ldots\ldots & \\ a_{r1}x_1 + \ldots + a_{rr}x_r &= b_r - a_{r,\,r+1}x_{r+1} - \ldots - a_{rm}x_m. \end{aligned}\right\} \tag{3}$$

Перенесённые направо неизвестные называются *свободными неизвестными,* так как мы можем придать каждому из них произвольное значение, подставить в систему (3) и решить её относительно неизвестных $x_1, \ldots, x_r$ (так как определитель этой системы отличен от нуля). Таким образом, возникает бесконечное множество решений системы (1), получаемых изменением значений свободных неизвестных.

Легко видеть, что *таким образом получаются все решения системы* (3), а следовательно, и системы (1). В самом деле, если $x_1', \ldots, x_r', x_{r+1}', \ldots, x_m'$ есть какое-либо решение системы (3), положим в системе (3): $x_{r+1} = x_{r+1}', \ldots, x_m = x_m'$. При этих условиях система (3) должна однозначно определить значения неизвестных $x_1, \ldots, x_r$, удовлетворяющие системе (2). Но такие значения нам уже известны: они равны $x_1', \ldots, x_r'$, следовательно (в силу однозначности!), именно их мы и получим, решая систему (3) по указанному выше правилу.

Резюмируя сказанное, получаем такой результат:

Теорема. *Если общее значение ранга матрицы системы* (1) *и расширенной матрицы той же системы меньше числа неизвестных, то система имеет бесконечное множество решений. Если же общее значение рангов указанных матриц равно числу неизвестных, то система* (1) *имеет единственное решение.*

Отметим один важный частный случай этого результата.

Система (1) называется *однородной, если все свободные члены уравнений равны нулю.* В случае однородной системы ранг расширенной матрицы никогда не может отличаться от ранга матрицы коэффициентов при неизвестных: добавление к системе векторов нулевого вектора не меняет ранга системы. Поэтому решение однородной системы должно всегда существовать (хотя бы одно). Такой результат ясен сразу: если мы в однородной системе уравнений положим значения всех неизвестных равными нулю, то мы удовлетворим системе. Это нулевое решение обычно не представляет интереса (хотя бы уже потому, что оно никак не связано с коэффициентами уравнений, а следовательно, и с постановкой задачи, приведшей к этой системе). Интересными являются только решения, отличные от нулевого. Сформулированная только что теорема позволяет высказать следующее условие существования таких решений.

Теорема. *Для того чтобы система однородных уравнений обладала решением, отличным от нулевого, необходимо и достаточно, чтобы ранг матрицы системы был мечьше числа неизвестных.*

В самом деле, в таком случае должно существовать бесконечное множество решений данной системы. Так как нулевое решение — только одно, то должны существовать и ненулевые решения.

Эта полезная теорема принимает особенно простую форму в случае системы n уравнений с n неизвестными:

Теорема. *Для того чтобы однородная система n уравнений с n неизвестными имела ненулевое решение, необходимо и достаточно, чтобы определитель этой системы равнялся нулю.*

Действительно, из матрицы коэффициентов при неизвестных можно составить в этом случае только один минор n-го порядка — определитель системы (в матрице всего n строк и n столбцов). Поэтому обращение единственного минора порядка n в нуль необходимо и достаточно для того, чтобы ранг матрицы был меньше n.

§ 16. Геометрическая интерпретация. Системы с тремя неизвестными

Используем теперь наш геометрический аппарат для того, чтобы получить обзор всей совокупности решений рассматриваемой системы линейных уравнений. Для этого будем рассматривать любую комбинацию значений неизвестных $x_1, x_2, \ldots, x_m$ как вектор m-мерного числового пространства. Если рассматриваемая комбинация будет решением системы уравнений, то мы будем говорить, что этот вектор есть решение данной системы.

Так как в общем случае отнюдь не каждый вектор будет решением интересующей нас системы уравнений, то решения будут заполнять собою только некоторую часть всего пространства. Наша задача будет состоять в том, чтобы охарактеризовать эту часть.

Начнём со случая однородной системы:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m &= 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m &= 0. \end{aligned}$$

Простой подстановкой легко убедиться в том, что если векторы $x' = (x'_1, x'_2, \ldots, x'_m)$ и $x'' = (x''_1, x''_2, \ldots, x''_m)$ (мы их пишем для удобства в виде строк) являются решениями этой системы, то векторы $x + y = (x'_1 + x''_1, \ldots, x'_m + x''_m)$ и $kx' = (kx'_1, kx'_2, \ldots, kx'_m)$ (при любом числовом множителе k) также будут решениями этой системы. Таким образом, совокупность решений содержит вместе с любым вектором и все его числовые кратные, а вместе с любыми двумя векторами — их сумму. Другими словами (ср. § 11):

Совокупность решений однородной системы уравнений всегда является подпространством m-мерного числового пространства, где m — число неизвестных в системе.

Легко определить также и размерность этого подпространства. Именно, имеет место

Теорема. *Размерность подпространства решений однородной системы уравнений с m неизвестными равна $m - r$, где r — ранг матрицы системы.*

В самом деле, пусть мы уже выбрали r независимых уравнений данной системы, как это было сделано в предыдущем параграфе,

и выбрали подходящим образом $m-r$ «свободных неизвестных» $x_{r+1}, \ldots, x_m$. Тогда все решения исходной системы мы получим, придавая свободным неизвестным произвольные значения и определяя остальные неизвестные из системы уравнений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1r}x_r &= -a_{1,r+1}x_{r+1} - \ldots - a_{1m}x_m, \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rr}x_r &= -a_{r,r+1}x_{r+1} - \ldots - a_{rm}x_m. \end{aligned}$$

При этом в силу доказанного в § 15 *каждой комбинации значений свободных неизвестных будет соответствовать в точности одна комбинация значений остальных неизвестных*, а следовательно, и одно единственное решение первоначальной системы.

Придадим теперь свободным неизвестным последовательно следующие комбинации значений (число их равно $m-r$):

$$\begin{array}{l} x_{r+1} = 1, \ 0, \ \ldots, \ 0, \\ x_{r+2} = 0, \ 1, \ \ldots, \ 0, \\ \cdot \quad\quad \cdot \quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \quad \cdot \quad\quad \cdot \\ x_m \ \ = 0, \ 0, \ \ldots, \ 1. \end{array}$$

Этим комбинациям значений будут соответствовать векторы

$$x' = \begin{pmatrix} x_1' \\ \cdot \\ \cdot \\ \cdot \\ x_r' \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad x'' = \begin{pmatrix} x_1'' \\ \cdot \\ \cdot \\ \cdot \\ x_r'' \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad \ldots, \quad x^{(m-r)} = \begin{pmatrix} x_1^{(m-r)} \\ \cdot \\ \cdot \\ \cdot \\ x_2^{(m-r)} \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \qquad (1)$$

являющиеся решениями первоначальной системы. Покажем, что они образуют базис подпространства решений. Тем самым будет установлено, что размерность этого подпространства равна $m-r$ (числу базисных векторов).

Для доказательства нужного утверждения достаточно обнаружить, что векторы (1) линейно независимы и что любой вектор, являющийся решением данной системы, представляется их линейной комбинацией.

Первое обстоятельство очевидно: линейная комбинация рассматриваемых векторов имеет вид

$$k_1x' + k_2x'' + \ldots + k_{m-r}x^{(m-r)} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ k_1 \\ k_2 \\ \cdot \\ \cdot \\ \cdot \\ k_{m-r} \end{pmatrix}, \qquad (2)$$

где в столбце точки означают r первых чисел. Очевидно, что эта линейная комбинация может оказаться нулевым столбцом только в том случае, когда все числа $k_1, k_2, \ldots, k_{m-r}$ равны нулю.

Пусть теперь имеется произвольное решение данной однородной системы

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_r \\ x_{r+1} \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{pmatrix}. \qquad (3)$$

Пользуясь выражением (2), легко видеть, что можно образовать линейную комбинацию решений (1), в которой свободные неизвестные будут иметь те же значения, что и в решении (3): для этого достаточно взять $k_1 = x_{r+1}$, $k_2 = x_{r+2}, \ldots, k_{m-r} = x_m$. Полученная так линейная комбинация решений (1) должна оказаться также решением заданной системы (ибо множество решений является подпространством). Замечая теперь, что у рассматриваемой системы может быть *только одно решение с данными значениями свободных неизвестных,* заключаем отсюда, что полученная линейная комбинация должна совпасть с решением (3), а это и нужно.

Полученный результат охватывает и случай, когда $r = m$, т. е. когда свободные неизвестные отсутствуют: в этом случае существует только одно нулевое решение, само по себе образующее подпространство, которое не содержит ни одного линейно независимого вектора и которое поэтому естественно называть *нульмерным.*

Теперь мы в состоянии рассмотреть случай произвольной неоднородной системы [система (1) предыдущего параграфа]. Условимся называть систему однородных уравнений, получаемую из данной системы приравниванием нулю её левых частей, *соответствующей* (данной системе) *однородной системой*. Для этой системы [система (1) этого параграфа] совокупность всех решений является некоторым подпространством m-мерного числового пространства.

Рассмотрим случай, когда данная система имеет хотя бы одно решение $x_{10}, x_{20}, \ldots, x_{m0}$. Имеет место следующий основной результат:

Теорема. *Совокупность решений неоднородной системы получается, если к каждому вектору подпространства решений соответствующей однородной системы прибавить* (одно и то же для всех векторов) *решение данной неоднородной системы уравнений.*

Для доказательства введём обозначение

$$x_1 = x_1' + x_{10}, \quad x_2 = x_2' + x_{20}, \quad \ldots, \quad x_m = x_m' + x_{m0},$$

где $x_1', x_2', \ldots, x_m'$ — новые неизвестные. Подставляя эти выражения $x_1, x_2, \ldots, x_m$ через $x_1', x_2', \ldots, x_m'$ в уравнения системы (1) § 15, получим систему равенств

$$\left.\begin{aligned} &a_{11}x_1' + a_{12}x_2' + \ldots + a_{1m}x_m' + \\ &\qquad + (a_{11}x_{10} + a_{12}x_{20} + \ldots + a_{1m}x_{m0}) = b_1, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{n1}x_1' + a_{n2}x_2' + \ldots + a_{nm}x_m' + \\ &\qquad + (a_{n1}x_{10} + a_{n2}x_{20} + \ldots + a_{nm}x_{m0}) = b_n. \end{aligned}\right\} \quad (4)$$

Так как $x_{10}, x_{20}, \ldots, x_{m0}$ есть решение исходной системы, то стоящие в скобках суммы в левых частях будут попросту равны правым частям уравнений. Поэтому система (4) совершенно равносильна такой:

$$\begin{aligned} &a_{11}x_1' + a_{12}x_2' + \ldots + a_{1m}x_m' = b_1, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{n1}x_1' + a_{n2}x_2' + \ldots + a_{nm}x_m' = b_n, \end{aligned}$$

т. е. оказывается *однородной системой, соответствующей данной системе.*

Проведённое преобразование показывает, что вектор $(x_1, x_2, \ldots, x_m)$ будет решением первоначальной системы в том и только в том случае, когда вектор $(x_1', x_2', \ldots, x_m')$ будет решением соответствующей однородной системы. Но это и есть требуемый результат.

Сказанное становится совершенно наглядным в случае систем уравнений с тремя неизвестными, так как в этом случае трёхмерные

числовые векторы можно себе представлять в виде обычных «геометрических» векторов, имеющих те же координаты (в какой-либо координатной системе), что и данные числовые векторы. При этом рассмотрении мы можем сразу оставить в системе только независимые уравнения, число которых, как всегда, равно рангу данной системы. Так как ранг не может превышать число неизвестных, то возможны только следующие четыре случая:

А) Ранг системы равен нулю. Независимых уравнений нет, т. е. все уравнения являются тождествами: коэффициенты при неизвестных и свободные члены равны нулю. В этом случае, очевидно, решения заполняют всё пространство. Связь между решениями неоднородной системы и решениями соответствующей однородной системы не имеет смысла рассматривать, так как возможны только однородные системы, удовлетворяющие поставленным условиям.

Б) Ранг системы равен единице. Независимое уравнение одно

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1. \tag{5}$$

Соответствующая однородная система также состоит из одного уравнения

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0. \tag{6}$$

В качестве свободных неизвестных можно выбрать любые два, но так, чтобы коэффициент при третьем неизвестном был отличен от нуля. Если таким коэффициентом является a_{11}, то можно переписать уравнение (5) в виде

$$x_1 = -\frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3.$$

Пространство решений для уравнения (6) — двумерное. Его базис мы получим, придавая x_2 и x_3 следующие комбинации значений: 1, 0 и 0, 1. Соответствующие решения однородного уравнения будут:

$$\left(-\frac{a_{12}}{a_{11}}, 1, 0\right) \quad \text{и} \quad \left(-\frac{a_{13}}{a_{11}}, 0, 1\right).$$

Если изобразить эти решения в виде векторов обычного трёхмерного пространства, то совокупность всех решений однородного уравнения представится множеством векторов плоскости L_2, «натянутой» на построенные так векторы.

Чтобы построить совокупность всех решений неоднородного уравнения, достаточно изобразить вектором x_0 одно из его решений. Тогда, согласно изложенному выше, прибавляя этот вектор ко всем векторам построенной только что плоскости, мы получим требуемое. Легко усмотреть, что концы всех построенных таким образом векторов (если считать, что их начала все расположены в начале координат) располагаются на плоскости, проведённой через конец вектора x_0 параллельно плоскости L_2.

В) Ранг системы равен двум. Независимых уравнений два. Пространство решений соответствующей однородной системы будет одномерное, т. е. будет являться совокупностью векторов, лежащих на некоторой прямой L_1, проходящей через начало координат. Если x_0 — любое решение неоднородной системы, то совокупность всех её решений геометрически представится множеством векторов, концы которых располагаются на прямой, проведённой через конец вектора x_0 параллельно прямой L_1.

Г) Ранг системы равен трём. В этом случае однородная система имеет только нулевое решение. Данная система будет иметь также только одно решение, так что «совокупность всех решений» представится одним единственным вектором.

Геометрический смысл этого становится ещё проще, если вместо векторов говорить о точках с теми же координатами, что и векторы, о которых шла речь. При этом становится привычным говорить о геометрическом месте решений. Содержание результатов, полученных в только что рассмотренных случаях, может быть высказано в этих терминах так:

Геометрическое место решений совместной системы линейных уравнений с тремя неизвестными будет всем пространством, если система тождественная, плоскостью, если система содержит лишь одно независимое уравнение, прямой, если в системе два независимых уравнения, наконец, точкой, если в данной системе имеется три независимых уравнения.

Читатель без труда узнает в этом факты, устанавливаемые в аналитической геометрии.

В заключение следует заметить, что всё предыдущее изложение можно было бы провести, всюду пользуясь понятием «точка» вместо понятия «вектор». Однако необходимость пользоваться такими алгебраическими операциями, как сложение, делает это не очень удобным: непривычность оборотов речи вроде «сложение точек» только затруднила бы читателя и не позволила бы ему глубже разглядеть геометрическое содержание излагаемых фактов.

§ 17. Применение к системам уравнений высших степеней

Последняя теорема § 15 играет во многих случаях исключительно важную роль и находит применение иногда в самых неожиданных вопросах. В качестве примера такого применения рассмотрим решение системы двух уравнений с двумя неизвестными в случае, когда степени этих уравнений совершенно произвольны.

Пусть дана такая система

$$\left.\begin{aligned} F(x, y) &= 0, \\ G(x, y) &= 0, \end{aligned}\right\} \tag{1}$$

где $F(x, y)$ и $G(x, y)$ — произвольно заданные многочлены от

неизвестных x и y. Требуется определить значения x и y так, чтобы они удовлетворяли обоим уравнениям.

Если бы значение одного из неизвестных было известно, то определение значения другого сводилось бы к решению уравнений уже с одним неизвестным. Для этого достаточно подставить известное значение в оба уравнения, решить их каждое отдельно и после этого посмотреть, какие значения второго неизвестного будут общими в обоих случаях. Таким образом, если заниматься только вопросом, специфическим для системы уравнений с двумя неизвестными, остаётся лишь найти те значения одного из неизвестных, при которых уравнения (1) могут иметь общие корни (относительно другого неизвестного).

Для этого перепишем уравнения (1), объединяя в них члены с одинаковыми степенями y и вынося эти степени за скобки в каждой группе членов. В скобках останутся многочлены от одного неизвестного x, и система (1) примет вид

$$\left.\begin{aligned} a_0(x)y^m + a_1(x)y^{m-1} + \ldots + a_m(x) = 0, \\ b_0(x)y^n + b_1(x)y^{n-1} + \ldots + b_n(x) = 0. \end{aligned}\right\} \qquad (2)$$

Пусть x_0, y_0 — одно из решений этой системы. Тогда будут иметь место равенства

$$a_0(x_0)y_0^m + a_1(x_0)y_0^{m-1} + \ldots + a_m(x_0) = 0$$

и

$$b_0(x_0)y_0^n + b_1(x_0)y_0^{n-1} + \ldots + b_n(x_0) = 0,$$

а вместе с ними и ряд равенств, получаемых из написанных умножением левых частей соответственно на $y_0^{n-1}, y_0^{n-2}, \ldots, y_0^0 = 1$ и на $y_0^{m-1}, y_0^{m-2}, \ldots, y_0^0 = 1$:

$$\left.\begin{aligned} a_0(x_0)y_0^{m+n-1} + a_1(x_0)y_0^{m+n-2} + \ldots + a_m(x_0)y_0^{n-1} = 0, \\ a_0(x_0)y_0^{m+n-2} + \ldots + a_m(x_0)y_0^{n-2} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0(x_0)y_0^m + \ldots + a_m(x_0) = 0, \\ b_0(x_0)y_0^{m+n-1} + b_1(x_0)y_0^{m+n-2} + \ldots + b_n(x_0)y^{m-1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_0(x_0)y_0^n + \ldots + b_n(x_0) = 0. \end{aligned}\right\} \qquad (3)$$

Заметим теперь, что число написанных равенств $m+n$ в точности совпадает с числом встречающихся в них степеней y_0 от нулевой до $(m+n-1)$-й включительно. Равенства (3) означают, что эти

степени составляют решение следующей однородной системы линейных уравнений:

$$\left.\begin{aligned} a_0(x_0)\,u_{m+n} + a_1(x_0)\,u_{m+n-1} + \ldots + a_m(x_0)\,u_n &= 0,\\ a_0(x_0)\,u_{m+n-1} + \ldots + a_m(x_0)\,u_{n-1} &= 0,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ a_0(x_0)\,u_{m+1} + \ldots + a_m(x_0)\,u_1 &= 0,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ b_0(x_0)\,u_{m+n} + b_1(x_0)\,u_{m+n-1} + \ldots + b_n(x_0)\,u_m &= 0,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ b_0(x_0)\,u_{n+1} + \ldots + b_n(x_0)\,u_1 &= 0. \end{aligned}\right\} \qquad (4)$$

Решение $u_1 = 1$, $u_2 = y_0, \ldots, u_{m+n} = y_0^{m+n-1}$ отлично от нулевого, так как значение неизвестного u_1 равно 1. Но для существования такого решения необходимо, чтобы определитель системы (4) был равен нулю, т. е.

$$\begin{vmatrix} a_0(x_0) & a_1(x_0) & \ldots & a_m(x_0) & 0 & \ldots & 0\\ 0 & a_0(x_0) & \ldots & \ldots & a_m(x_0) & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \ldots & a_0(x_0) & \ldots & \ldots & a_m(x_0)\\ b_0(x_0) & b_1(x_0) & \ldots & \ldots & b_n(x_0) & 0 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \ldots & b_0(x_0) & \ldots & \ldots & b_n(x_0) \end{vmatrix} = 0. \qquad (5)$$

Результат справедлив, какое бы решение x_0, y_0 данной системы ни было взято, а это даёт следующий путь отыскания *всех* значений $x = x_0$, которые могут входить в решения системы (1): по данным уравнениям системы составляем уравнение

$$\left.\begin{matrix} n \text{ строк} \\ \\ \\ m \text{ строк} \\ \\ \\ \end{matrix}\right| \begin{matrix} a_0(x) & a_1(x) & \ldots & a_m(x) & 0 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ & & & a_0(x) & \ldots & \ldots & a_m(x)\\ b_0(x) & b_1(x) & \ldots & b_n(x) & 0 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \ldots & b_0(x) & \ldots & \ldots & b_n(x) \end{matrix} = 0. \qquad (6)$$

В силу доказанного любое значение x_0, входящее в какое-либо решение системы (1), будет его корнем. Поэтому дальнейшее сводится к тому, что мы находим все возможные корни уравнения (6), подставляем их последовательно в систему (1) и находим, какие общие корни (для неизвестного y) они при этом имеют. Таким образом, задача сведена к решению систем уравнений с одним

неизвестным и конечному числу подстановок различных значений неизвестного x.

Следует, впрочем, отметить, что равенство (5) является лишь необходимым условием для того, чтобы значение $x=x_0$ входило в решение системы (1): хотя при любом значении x, удовлетворяющем уравнению (6), система однородных уравнений (4) и будет иметь решение, нельзя ручаться, что при этом значения неизвестных будут соответствующими степенями одного и того же числа. Более детальное исследование показывает, однако, что «лишними» могут оказаться только те значения x, при которых оба коэффициента $a_0(x)$ и $b_0(x)$ в уравнениях (2) обращаются в нуль одновременно.

Определитель, стоящий в левой части уравнения (6), называется *результантом* данной системы уравнений.

§ 18. Дополнительные замечания

1. Эквивалентность систем линейных уравнений. Две системы уравнений (в частности, линейных) называются *эквивалентными*, если каждое решение одной из них является также решением другой, и наоборот.

В случае систем линейных уравнений имеет место весьма простое соотношение между уравнениями эквивалентных систем:

Если две системы линейных уравнений эквивалентны, то каждое из уравнений любой из этих систем получается из уравнений другой системы умножением обеих частей каждого из них на некоторое число и последующим сложением.

Для краткости в таких случаях говорят, что уравнение является линейной комбинацией уравнений системы.

Доказательство этого можно легко получить из основной теоремы о совместности систем. В самом деле, если две системы эквивалентны, то соединением их получим снова систему, эквивалентную обеим данным. У новой системы будет то же самое общее решение, а следовательно, и то же самое число свободных неизвестных. Но это означает, что при таком соединении систем ранг расширенной матрицы не изменяется. Другими словами, все уравнения наших систем будут линейными комбинациями уравнений некоторой максимальной линейно независимой системы. Так как эту последнюю всегда можно выбрать из уравнений одной из данных систем (опять-таки в силу равенства рангов), то наше утверждение доказано.

2. О вычислении ранга матрицы. Способ вычисления ранга матрицы, основанный на самом определении ранга, весьма утомителен: приходится вычислять очень большое число миноров. Трудность уменьшается, если в рассматриваемой матрице значительное число элементов обращается в нуль, так как при этом равенство многих

миноров нулю становится очевидным, и остаётся провести вычисление лишь относительно небольшого числа определителей.

Следующее замечание позволяет искусственно упрощать рассматриваемую матрицу, если это оказывается необходимым:

Если к одному из векторов системы $e'_1, \ldots, e'_m$ *прибавить любую линейную комбинацию других векторов той же системы, то ранг системы не изменится.*

В самом деле, рассмотрим системы

$$e'_1, \ldots, e'_i, \ldots, e'_m \text{ и } e'_1, \ldots, e'_i + \sum_{j \neq i} k_j e'_j, \ldots, e'_m.$$

По построению вторая из этих систем оказывается линейно зависящей от первой. Столь же очевидно, что векторы первоначальной системы могут быть выражены линейными комбинациями векторов второй системы; мы видим, таким образом, что обе системы эквивалентны, а следовательно, они имеют один и тот же ранг.

Переводя содержание сделанного замечания на язык матриц и учитывая, что в этом случае «векторами» можно называть по желанию строки или столбцы данной матрицы, получаем теорему:

Теорема. *Ранг матрицы не изменится, если к одному из её столбцов (или к одной из строк) прибавить линейную комбинацию других столбцов (строк) той же матрицы.*

Следующий пример показывает, насколько эффективным оказывается иногда пользование этим простым предложением:

Пусть нужно вычислить ранг матрицы

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}. \tag{1}$$

Если из каждой строки матрицы вычесть предыдущую строку (это, как легко видеть, не изменит ранга), то получим матрицу

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 \end{pmatrix}. \tag{2}$$

Здесь уже видно, что любой минор третьего порядка будет содержать хотя бы две одинаковые строки и поэтому обращается в нуль. А так как отличный от нуля минор второго порядка в матрице (2) существует (например, минор $\begin{vmatrix} 1 & 2 \\ 5 & 5 \end{vmatrix}$), то ранг данной матрицы равен 2.

3. О существовании решений систем уравнений в различных числовых полях. Одной из причин, вызвавших появление в математической теории различных числовых (и не только числовых) полей, являлось то обстоятельство, что существуют уравнения, которые в одних полях имеют решения, а в других — нет. Так, например, известно, что уравнение $x^2 + 1 = 0$, коэффициенты которого принадлежат полю действительных чисел, не имеет решений в этом поле, а в поле комплексных чисел у того же уравнения имеются решения i и $-i$. Причиной, заставившей рассматривать поле комплексных чисел, было именно желание добиться того, чтобы всякое алгебраическое уравнение (первоначально с действительными коэффициентами) имело решения.

Подобное появление решений при расширении поля не может иметь места в случае линейных уравнений:

Если система линейных уравнений с коэффициентами из поля K не имеет решений в этом поле, то она не может иметь решений ни в каком другом, более широком, поле.

В самом деле, существование или несуществование решений связано с соотношением между рангом матрицы системы и расширенной матрицы. Но эти ранги не могут измениться при расширении рассматриваемого числового поля, так как не меняются даже значения миноров матриц (ибо вычисление определителей сводится к действиям сложения, вычитания и деления).

Это обстоятельство объясняет, почему вопрос о расширении поля чисел не мог возникнуть исторически раньше, чем начали заниматься задачами, приводящими к уравнениям второй и более высоких степеней.

ГЛАВА III

ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ ПЛОСКОСТИ И ТРЁХМЕРНОГО ПРОСТРАНСТВА

§ 19. Метрика. Скалярное произведение векторов

На общие векторные пространства, являющиеся естественным обобщением векторного пространства элементарной геометрии, можно распространить ряд важнейших свойств последнего. Сюда относятся, прежде всего, метрические свойства, связанные с возможностью измерения длин отрезков и величин углов.

Понятия длины вектора и угла можно ввести в случае векторных пространств любой размерности над любым числовым полем. Однако мы предпочитаем ограничиться в дальнейшем изложении рассмотрением обычного трёхмерного пространства и обычной плоскости. При этом в качестве основного поля, над которым определены наши пространства, принимается поле действительных чисел.

Для того чтобы связать длину вектора и угол между двумя векторами с выражением векторов через их координаты, удобнее всего ввести понятие о так называемом скалярном произведении векторов.

Скалярным произведением двух данных векторов называется произведение их длин, умноженное на косинус угла между ними.

Скалярное произведение векторов $\boldsymbol{a}$ и $\boldsymbol{b}$ мы будем обозначать $(\boldsymbol{a}, \boldsymbol{b})$. Удобство применения скалярного произведения при рассмотрении метрических свойств пространства обусловлено тем, что через него можно выразить как длину вектора, так и угол между двумя векторами. Если обозначить длину вектора обычным знаком абсолютной величины, то из выражения скалярного произведения

$$(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cos \alpha \tag{1}$$

легко видеть, что имеют место формулы:

$$|\boldsymbol{a}|^2 = (\boldsymbol{a}, \boldsymbol{a}), \quad \cos \alpha = \frac{(\boldsymbol{a}, \boldsymbol{b})}{\sqrt{(\boldsymbol{a}, \boldsymbol{a})(\boldsymbol{b}, \boldsymbol{b})}}, \tag{2}$$

сводящие вычисление длин и углов к вычислению скалярных произведений.

После этих замечаний легко получить аналитическое выражение скалярного произведения векторов через их координаты. Рассмотрим сначала выражение длины вектора. При этом в качестве базиса пространства удобно принять систему трёх (а в случае плоскости — двух) взаимно перпендикулярных векторов e_1, e_2, e_3, длина каждого из которых равна единице. В таком случае координатами произвольного вектора x будут взятые с надлежащими знаками длины отрезков OX_1, OX_2 и OX_3, служащих проекциями вектора x

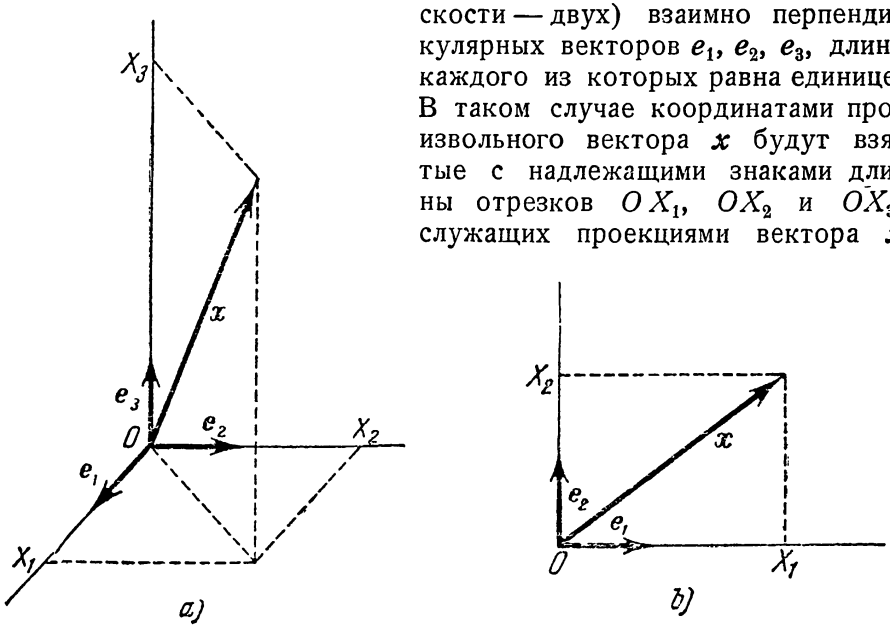

Рис. 6.

на направления векторов e_1, e_2, e_3 (рис. 6). Поэтому в случае, если $x = e_1x_1 + e_2x_2 + e_3x_3$, длина вектора x определяется (как длина диагонали прямоугольного параллелепипеда) формулой

$$|x|^2 = x_1^2 + x_2^2 + x_3^2. \tag{3}$$

(В случае плоскости получается формула $|x|^2 = x_1^2 + x_2^2$, отличающаяся от предыдущей только отсутствием третьего члена.)

Чтобы получить координатное выражение скалярного произведения двух векторов $x = e_1x_1 + e_2x_2 + e_3x_3$ и $y = e_1y_1 + e_2y_2 + e_3y_3$, рассмотрим их сумму $x + y = e_1(x_1 + y_1) + e_2(x_2 + y_2) + e_3(x_3 + y_3)$. Длина вектора $x + y$ может быть выражена двумя способами: с одной стороны, по формуле (3) имеем:

$$|x + y|^2 = (x_1 + y_1)^2 + (x_2 + y_2)^2 + (x_3 + y_3)^2 =$$
$$= x_1^2 + x_2^2 + x_3^2 + y_1^2 + y_2^2 + y_3^2 + 2(x_1y_1 + x_2y_2 + x_3y_3);$$

с другой стороны (рис. 7), квадрат длины диагонали параллелограмма равен

$$|x+y|^2=|x|^2+|y|^2+2|x|\cdot|y|\cos\alpha=$$
$$=x_1^2+x_2^2+x_3^2+y_1^2+y_2^2+y_3^2+2(x,\ y).$$

Сравнивая оба полученных выражения, видим, что

$$(x,\ y)=x_1y_1+x_2y_2+x_3y_3. \qquad (4)$$

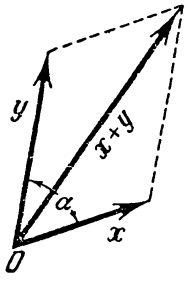

Рис. 7.

Таково выражение скалярного произведения в избранной нами системе координат.

Выражение (4) позволяет непосредственно усмотреть ряд свойств скалярного произведения, часть из которых, впрочем, легко усматривается также и из самого его определения.

1. *Скалярное произведение не зависит от порядка сомножителей.*

2. *Числовой множитель можно выносить из-под знака скалярного произведения*: $(k\boldsymbol{a},\ \boldsymbol{b})=k(\boldsymbol{a},\ \boldsymbol{b})$, каковы бы ни были векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ и число k.

3. *Для скалярного произведения имеет место распределительный закон*:

$$(\boldsymbol{a},\ \boldsymbol{b}+\boldsymbol{c})=(\boldsymbol{a},\ \boldsymbol{b})+(\boldsymbol{a},\ \boldsymbol{c}). \qquad (5)$$

Доказательство всех этих свойств можно провести прямой проверкой, пользуясь полученным выражением скалярного произведения. Например, в случае третьего свойства эта проверка происходит так: пусть $\boldsymbol{a}=\boldsymbol{e}_1a_1+\boldsymbol{e}_2a_2+\boldsymbol{e}_3a_3$, $\boldsymbol{b}=\boldsymbol{e}_1b_1+\boldsymbol{e}_2b_2+\boldsymbol{e}_3b_3$, $\boldsymbol{c}=\boldsymbol{e}_1c_1+\boldsymbol{e}_2c_2+\boldsymbol{e}_3c_3$. Тогда

$$\boldsymbol{b}+\boldsymbol{c}=\boldsymbol{e}_1(b_1+c_1)+\boldsymbol{e}_2(b_2+c_2)+\boldsymbol{e}_3(b_3+c_3),$$

а значит, в силу формулы (4)

$$(\boldsymbol{a},\ \boldsymbol{b}+\boldsymbol{c})=a_1(b_1+c_1)+a_2(b_2+c_2)+a_3(b_3+c_3).$$

С другой стороны, по той же формуле (4)

$$(\boldsymbol{a},\ \boldsymbol{b})=a_1b_1+a_2b_2+a_3b_3;\quad (\boldsymbol{a},\ \boldsymbol{c})=a_1c_1+a_2c_2+a_3c_3,$$

откуда

$$(\boldsymbol{a},\ \boldsymbol{b})+(\boldsymbol{a},\ \boldsymbol{c})=a_1(b_1+c_1)+a_2(b_2+c_2)+a_3(b_3+c_3).$$

Совпадение полученных выражений доказывает равенство (5). Проверка остальных свойств проходит ещё проще, и мы оставляем её читателю.

Распределительный закон применим также и в случае сумм, состоящих из многих слагаемых, причём доказательство этого не требует нового обращения к формуле (4), а может быть проведено

на основе уже доказанных свойств. Способ, которым всё сводится к свойствам 1 и 3, — такой же, как в случае элементарной алгебры. Он яснее всего виден на таком примере:

$$(a+b,\ c+d)=(a+b,\ c)+(a+b,\ d)=(c,\ a+b)+(d,\ a+b)=$$
$$=(c,\ a)+(c,\ b)+(d,\ a)+(d,\ b)=$$
$$=(a,\ c)+(b,\ c)+(a,\ d)+(b,\ d),$$

в котором произведение $(a+b,\ c+d)$ сначала рассматривается как произведения одного вектора $a+b$ на сумму векторов c и d, затем используется возможность перестановки множителей, снова применяется формула (5) и, наконец, делается обратная перестановка множителей. Доказательство в случае любых сумм проводится индукцией.

Свойства 1—3 позволяют получить выражение скалярного произведения в любой системе координат, т. е. в случае базиса, состоящего из трёх произвольных векторов e_1', e_2', e_3', не лежащих в одной плоскости: если

$$x=e_1'x_1'+e_2'x_2'+e_3'x_3',\quad y=e_1'y_1'+e_2'y_2'+e_3'y_3',$$

то

$$(x,\ y)=(e_1',\ e_1')\,x_1'y_1'+(e_1',\ e_2')\,x_1'y_2'+(e_2',\ e_1')\,x_2'y_1'+$$
$$+(e_1',\ e_3')\,x_1'y_3'+(e_3',\ e_1')\,x_3'y_1'+(e_2',\ e_2')\,x_2'y_2'+$$
$$+(e_2',\ e_3')\,x_2'y_3'+(e_3',\ e_2')\,x_3'y_2'+(e_3',\ e_3')\,x_3'y_3'. \quad (6)$$

Выражение, стоящее в правой части равенства (6), замечательно тем, что в каждый его член входит одна из координат каждого из рассматриваемых векторов x и y, причём точно в первой степени. Подобного рода выражения называются *билинейными формами* от координат x_1', x_2', x_3' и y_1', y_2', y_3'. В случае, если векторы e_1', e_2', e_3' взаимно перпендикулярны, то все скалярные произведения любого из них на остальные обращаются в нуль (косинус угла между множителями в этом случае равен нулю), и формула (6) принимает вид

$$(x,\ y)=(e_1',\ e_1')\,x_1'y_1'+(e_2',\ e_2')\,x_2'y_2'+(e_3',\ e_3')\,x_3'y_3'. \quad (7)$$

В случае, когда, кроме того, длина каждого из векторов e_1', e_2', e_3' равна единице, формула ещё более упрощается и обращается в такую:

$$(x,\ y)=x_1'y_1'+x_2'y_2'+x_3'y_3', \quad (8)$$

как, впрочем, и следовало ожидать, ибо в сущности именно от этой формулы мы и отправлялись.

Сравнение формул (6), (7) и (8) делает очевидным, что при решении любых метрических задач, т. е. задач, связанных с измерением или вычислением любых длин и углов, наиболее удобно

пользоваться базисом, состоящим из взаимно перпендикулярных единичных векторов. Такие базисы (или системы координат) называются *ортонормальными*.

Не повторяя для плоскости всех приведённых рассуждений, ограничимся тем, что приведём формулы, соответствующие формулам (4), (6) и (8):

$$\left.\begin{aligned}&(x,\ y)=x_1y_1+x_2y_2,\\&(x,\ y)=(e_1',\ e_1')\,x_1'y_1'+(e_1',\ e_2')\,x_1'y_2'+(e_2',\ e_1')\,x_2'y_1'+(e_2',\ e_2')\,x_2'y_2',\\&(x,\ y)=x_1'y_1'+x_2'y_2'.\end{aligned}\right\}\quad(4')$$

Последняя формула даёт выражение скалярного произведения в любом ортонормальном базисе плоскости.

§ 20. Преобразование координат

Хотя, как было только что отмечено, при решении метрических задач всегда оказывается наиболее удобным пользование ортонормальным базисом, довольно часто возникает необходимость в процессе решения задачи изменить координатную систему. При этом изменяются также и координаты векторов. Задача преобразования координат состоит в установлении связи между координатами произвольного, но одного и того же вектора относительно различных базисов пространства.

Эта задача решается совершенно одинаково в обоих интересующих нас случаях — плоскости и трёхмерного пространства: отличие состоит лишь в числе векторов базиса и числе координат. Сначала получим явные формулы в случае пространства.

Пусть задан произвольный базис e_1, e_2, e_3, относительно которого координаты некоторого произвольно заданного вектора x будут: x_1, x_2, x_3, т. е. $x=e_1x_1+e_2x_2+e_3x_3$. Если e_1', e_2', e_3' — некоторый другой базис пространства и x_1', x_2', x_3' — координаты вектора x по отношению к этому базису, то для установления зависимостей между «старыми» и «новыми» координатами можно поступить следующим образом: векторы e_1', e_2', e_3', как и векторы пространства, однозначно представляются линейными комбинациями векторов e_1, e_2, e_3 «старого» базиса. Удобно коэффициенты этих линейных комбинаций, как это мы часто делали, обозначить одной и той же буквой и различать только индексами, характеризующими роль каждого из коэффициентов. Таким образом, соотношения между векторами e_1', e_2', e_3' и e_1, e_2, e_3 запишутся в виде

$$\left.\begin{aligned}e_1'&=e_1c_{11}+e_2c_{21}+e_3c_{31},\\e_2'&=e_1c_{12}+e_2c_{22}+e_3c_{32},\\e_3'&=e_1c_{13}+e_2c_{23}+e_3c_{33},\end{aligned}\right\}\quad(1)$$

(первый индекс соответствует вектору «старого» базиса, а второй — вектору «нового»). Матрица

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}, \tag{2}$$

составленная из коэффициентов формул (1), уже была названа раньше (см. § 13) *матрицей перехода от базиса* e_1, e_2, e_3 *к базису* e'_1, e'_2, e'_3. В силу линейной независимости векторов e'_1, e'_2, e'_3 её определитель должен быть отличным от нуля. Подставим теперь выражения (1) в выражение вектора x через новый базис. Тогда получим:

$$x = e'_1x'_1 + e'_2x'_2 + e'_3x'_3 = e_1(c_{11}x'_1 + c_{12}x'_2 + c_{13}x'_3) + \\ + e_2(c_{21}x'_1 + c_{22}x'_2 + c_{23}x'_3) + e_3(c_{31}x'_1 + c_{32}x'_2 + c_{33}x'_3). \tag{3}$$

Последнее из выражений представляет собою не что иное, как представление вектора x линейной комбинацией векторов e_1, e_2, e_3 старого базиса. Но так как любой вектор представляется линейной комбинацией векторов базиса т о л ь к о о д н и м способом, то выражение (2) должно совпадать с исходным выражением $x = e_1x_1 + e_2x_2 + e_3x_3$ вектора x. Таким образом, должны выполняться равенства

$$\left.\begin{aligned} x_1 &= c_{11}x'_1 + c_{12}x'_2 + c_{13}x'_3, \\ x_2 &= c_{21}x'_1 + c_{22}x'_2 + c_{23}x'_3, \\ x_3 &= c_{31}x'_1 + c_{32}x'_2 + c_{33}x'_3, \end{aligned}\right\} \tag{4}$$

которые и связывают «старые» координаты x_1, x_2, x_3 вектора x с «новыми» координатами x'_1, x'_2, x'_3. Из равенств (4) можно получить однозначно выражения новых координат через старые, так как определитель матрицы (2) отличен от нуля. Обратим внимание на то, что коэффициентами в выражениях (4) служат те же элементы матрицы перехода, что и в выражениях (1) векторов нового базиса. Ниже мы увидим, что это даёт возможность ввести очень простую и наглядную символическую запись формул (1) и (4).

Обращаясь к случаю плоскости, на которой заданы два различных базиса e_1, e_2 и e'_1, e'_2, связанных соотношениями

$$\left.\begin{aligned} e'_1 &= e_1c_{11} + e_2c_{21}, \\ e'_2 &= e_1c_{12} + e_2c_{22} \end{aligned}\right\} \tag{1'}$$

с матрицей перехода

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}, \tag{2'}$$

легко найдём, что координаты вектора $x = e_1x_1 + e_2x_2 = e'_1x'_1 + e'_2x'_2$ связаны соотношениями

$$\left.\begin{aligned} x_1 &= c_{11}x'_1 + c_{12}x'_2, \\ x_2 &= c_{21}x'_1 + c_{22}x'_2. \end{aligned}\right\} \tag{4'}$$

В § 19 уже было отмечено, что при решении метрических задач особенно удобны ортонормальные базисы. Поэтому важно заметить особенности строения формул преобразования координат *при переходе от одного ортонормального базиса к другому*. Начнём со случая плоскости. Если векторы e_1, e_2 составляют ортонормальный базис (т. е. взаимно перпендикулярны и имеют длины, равные единице), то скалярные произведения векторов нового базиса e'_1, e'_2 будут выражаться по формуле (4') § 19:

$$(e'_1, e'_1) = c_{11}^2 + c_{21}^2, \quad (e'_1, e'_2) = c_{11}c_{12} + c_{21}c_{22}, \quad (e'_2, e'_2) = c_{12}^2 + c_{22}^2.$$

Поэтому для того, чтобы векторы e'_1, e'_2 нового базиса были единичными и взаимно перпендикулярными, необходимо и достаточно выполнение соотношений

$$c_{11}^2 + c_{21}^2 = 1, \quad c_{11}c_{12} + c_{21}c_{22} = 0, \quad c_{12}^2 + c_{22}^2 = 1 \tag{5'}$$

между элементами матрицы перехода (2').

Таким же образом в случае пространства получаем соотношения

$$\left.\begin{aligned} c_{11}^2 + c_{21}^2 + c_{31}^2 &= 1, \quad c_{11}c_{12} + c_{21}c_{22} + c_{31}c_{32} = 0, \\ c_{12}^2 + c_{22}^2 + c_{32}^2 &= 1, \quad c_{11}c_{13} + c_{21}c_{23} + c_{31}c_{33} = 0, \\ c_{13}^2 + c_{23}^2 + c_{33}^2 &= 1, \quad c_{12}c_{13} + c_{22}c_{23} + c_{32}c_{33} = 0 \end{aligned}\right\} \tag{5}$$

между элементами матрицы перехода от одного ортонормального базиса к другому.

Преобразования координат, при которых совершается переход от одного ортонормального базиса к другому, называются *ортогональными преобразованиями*. Матрица перехода, соответствующая такому преобразованию, называется *ортогональной матрицей*.

Ортогональные преобразования особенно просты в случае плоскости: если вектор e'_1 нового ортогонального базиса образует с вектором e_1 старого базиса угол φ, то вектор e'_2 должен образовать с вектором e_1 один из двух углов $\frac{\pi}{2} + \varphi$ или $-\frac{\pi}{2} + \varphi$ (рис. 8). Рассматривая проекции векторов e'_1 и e'_2 на направления векторов e_1 и e_2, легко усмотреть, что векторы нового базиса будут выражаться в этих двух случаях так:

$$\left.\begin{aligned} e'_1 &= e_1 \cos\varphi + e_2 \sin\varphi, \\ e'_2 &= -e_1 \sin\varphi + e_2 \cos\varphi \end{aligned}\right\} \tag{6}$$

и

$$\left.\begin{aligned} e_1' &= e_1 \cos\varphi + e_2 \sin\varphi, \\ e_2' &= e_1 \sin\varphi - e_2 \cos\varphi, \end{aligned}\right\} \tag{6'}$$

а матрицы перехода будут:

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}. \tag{7}$$

Геометрическое различие между двумя отмеченными случаями состоит в том, что в первом из них базис e_1', e_2' может быть полу-

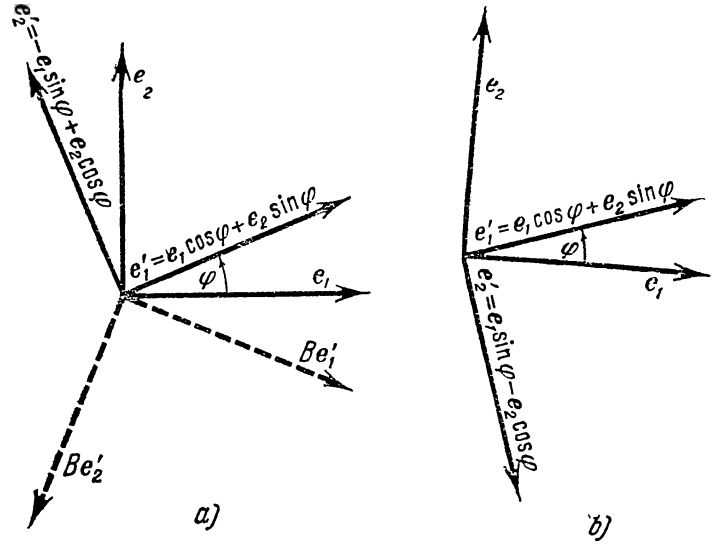

Рис. 8.

чен из базиса e_1, e_2 вращением последнего (в плоскости), а во втором базис e_1, e_2 никаким вращением не может быть переведён в базис e_1', e_2': если повернуть плоскость так, чтобы вектор e_1 совпал с вектором e_1', то после этого поворота вектор e_2 окажется не совпадающим, а противоположным вектору e_2' нового базиса. Это геометрическое различие аналитически выражается в том, что определитель матрицы перехода (7) в первом случае оказывается равным $+1$, а во втором случае — равным -1 [1]).

Аналогичные два класса ортогональных преобразований можно заметить и в случае пространства: геометрически ясно, что вращением пространства можно привести вектор e_1 к совпадению с вектором e_1'; после этого можно продолжать вращение, сохраняя

[1]) Достаточно убедиться в этом простым подсчётом: например,

$$\begin{vmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{vmatrix} = \cos^2\varphi + \sin^2\varphi = 1.$$

постоянной ось e_1', и заставить вектор e_2 совпасть с вектором e_2'. Что касается вектора e_3, то он после указанных вращений либо совпадает с вектором e_3', либо окажется противоположным ему. В последнем случае уже никаким поворотом векторы e_1, e_2, e_3 нельзя заставить совпасть с соответствующими векторами e_1', e_2' и e_3'. Аналитическое различие между этими двумя случаями оказывается таким же, как и для плоскости, но подметить это значительно труднее. Мы займёмся ещё этим вопросом после того, как подготовим в следующем параграфе нужный вспомогательный аппарат.

§ 21. Операции над матрицами

Многие соотношения, в частности те, которыми мы занимались в предыдущем параграфе, приобретают особенно отчетливый вид, если воспользоваться некоторыми формальными правилами действий с матрицами.

В предыдущей главе уже было определено сложение матриц, а также умножение матрицы на число. Эти определения оказались такими, что совокупность всех матриц, имеющих данное число строк и данное число столбцов, образует векторное пространство. Мы введём ещё две операции над матрицами. Первая из них, называемая *транспонированием*, по существу нам уже встречалась. Если дана матрица

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

то *транспонированной матрицей* для данной матрицы называется матрица

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix},$$

столбцами которой являются соответствующие строки данной матрицы. Мы часто будем обозначать матрицу не таблицей, как это сделано только что, а одной буквой (для этого будут всегда использоваться большие латинские буквы). В таком случае матрица, транспонированная для матрицы A, будет обозначаться значком T сверху: A^T. В частности, если данная матрица состоит всего из одного столбца

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

то транспонированная матрица $X^T=(x_1, x_2, \ldots, x_n)$ состоит из одной строки.

Кроме транспонирования, мы определим ещё умножение матриц. Определение умножения подсказывается видом формул (1), (4), (4′) предыдущего параграфа. Оно гласит:

Если заданы две матрицы A и B, причём число столбцов первой из них равно числу строк второй, то произведением их называется матрица AB, число строк которой совпадает с числом строк матрицы A, а число столбцов — с числом столбцов матрицы B, и такая, что на пересечении любой i-й строки и j-го столбца матрицы AB стоит сумма произведений соответствующих элементов i-й строки матрицы A и j-го столбца матрицы B.

В случае, если число столбцов матрицы A не равно числу строк матрицы B, произведение не определено. Из приведённого определения следует, в частности, что произведением одной строки на любую матрицу будет являться строчка, а произведением любой матрицы A на матрицу, состоящую из одного столбца, будет являться также матрица, состоящая из одного столбца. Ещё более частный случай — умножение одной строчки на столбец — даёт в результате матрицу, состоящую лишь из одного элемента.

Целесообразность такого определения умножения матриц будет видна дальше. Сейчас же мы ограничимся тем, что на нескольких примерах поясним высказанное определение. Прежде всего формулы (4) и (4′) предыдущего параграфа записываются в виде

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix},$$

а если для матриц перехода (2), (2′) ввести сокращённое обозначение одной буквой C, то эти формулы запишутся ещё короче и единообразней:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = C \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = C \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}.$$

Читателю предоставляется самому проверить, что ниженаписанные произведения образованы согласно сформулированному определению:

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 0 & 3 \end{pmatrix}; \quad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & 3 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Последний пример показывает, что квадратные матрицы вида

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

«ведут себя» подобно числу единица: если умножение на них возможно (т. е. если выполнено требование о числе строк и столбцов множителей), то от умножения на такую матрицу умножаемая матрица не изменяется. Однако первый пример показывает, что один из основных законов действий — *коммутативность или переместительность умножения — не имеет места для умножения матриц.*

Тем не менее оперировать с матрицами оказывается почти так же удобно, как с обычными числами, ибо *два других закона* операций — *сочетательный закон умножения* и *распределительный закон*, связывающий сложение с умножением, — остаются справедливыми при введённых нами определениях операций. Иными словами, каковы бы ни были матрицы A, B и C, имеют место соотношения

$$(A+B)C=AC+BC;\quad A(B+C)=AB+AC;\quad (AB)C=A(BC). \tag{1}$$

Убедиться в этом можно прямой проверкой, вычисляя отдельно правые и левые части написанных равенств. Например, если матрицы A, B и C равны, соответственно,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},\quad \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \quad \text{и} \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

то

$$(A+B)C=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}=$$

$$=\begin{pmatrix} (a_{11}+b_{11})c_1+(a_{12}+b_{12})c_2 \\ (a_{21}+b_{21})c_1+(a_{22}+b_{22})c_2 \end{pmatrix},$$

а

$$AC+BC=\begin{pmatrix} a_{11}c_1+a_{12}c_2 \\ a_{21}c_1+a_{22}c_2 \end{pmatrix}+\begin{pmatrix} b_{11}c_1+b_{12}c_2 \\ b_{21}c_1+b_{22}c_2 \end{pmatrix}=$$

$$=\begin{pmatrix} a_{11}c_1+a_{12}c_2+b_{11}c_1+b_{12}c_2 \\ a_{21}c_1+a_{22}c_2+b_{21}c_1+b_{22}c_2 \end{pmatrix}.$$

Совпадение результатов показывает справедливость в данном случае равенства $(A+B)C=AC+BC$. В общем случае проверка имеет тот же характер, но усложняется и запись делается более громоздкой. Для сочетательного закона $(AB)C=A(BC)$ мы проведём

общее доказательство в силу его меньшей очевидности и той роли, которую он играет во всём дальнейшем. Пусть матрицы A, B и C равны, соответственно,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{pmatrix}, \quad \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}, \quad \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1q} \\ c_{21} & c_{22} & \dots & c_{2q} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nq} \end{pmatrix}.$$

(Напомним ещё раз, что для возможности умножения необходимо, чтобы число столбцов первой матрицы равнялось числу строк второй, а число столбцов второй — числу строк третьей.) Тогда на пересечении любой i-й строки и j-го столбца произведения AB будет стоять число

$$d_{ij} = \sum_{\alpha=1}^{m} a_{i\alpha} b_{\alpha j},$$

на пересечении же i-й строки и k-го столбца произведения $(AB)C$ — число

$$f_{ik} = \sum_{\beta=1}^{n} d_{i\beta} c_{\beta k} = \sum_{\beta=1}^{n} \sum_{\alpha=1}^{m} a_{i\alpha} b_{\alpha\beta} c_{\beta k}.$$

С другой стороны, на пересечении j-й строки и k-го столбца произведения BC будет стоять:

$$d'_{jk} = \sum_{\beta=1}^{n} b_{j\beta} c_{\beta k},$$

а следовательно на пересечении i-й строки и k-го столбца произведения $A(BC)$ будет расположено число

$$f'_{ik} = \sum_{\alpha=1}^{m} a_{i\alpha} d'_{\alpha k} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} a_{i\alpha} b_{\alpha\beta} c_{\beta k}.$$

Так как это выражение совпадает с числом, стоящим на пересечении i-й строки и k-го столбца произведения $(AB)C$, то будет: $(AB)C = A(BC)$, что и нужно.

Весьма полезно заметить влияние *операции транспонирования на сумму и произведение матриц*. Правила транспонирования сумм и произведений имеют вид

$$(A+B)^T = A^T + B^T \quad \text{и} \quad (AB)^T = B^T A^T,$$

или словами: а) *транспонированная матрица для суммы матриц равна сумме транспонированных слагаемых*; б) *транспонированная матрица для произведения равна произведению транспонированных сомножителей в обратном порядке.*

Первое из этих правил очевидно. Что касается второго, то оно легко получается из определения умножения. Мы ограничимся проверкой его справедливости в двух частных случаях:

$$\text{a)}\quad (a_1,\ a_2)\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = (a_1b_1 + a_2b_2);\quad (b_1,\ b_2)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = (a_1b_1 + a_2b_2);$$

$$\text{b)}\quad (a_1,\ b_2)\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = (a_1b_{11} + a_2b_{21},\ a_1b_{12} + a_2b_{22});$$

$$\begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1b_{11} + a_2b_{21} \\ a_1b_{12} + a_2b_{22} \end{pmatrix}.$$

Доказательство в общем случае может быть проведено подобно тому, как это было сделано выше для сочетательного закона.

Умножение квадратных матриц одного и того же порядка (т. е. с одним и тем же числом строк и столбцов) замечательным образом связано с определителями этих матриц. Имеет место такая

Теорема. *Определитель произведения двух квадратных матриц равен произведению определителей сомножителей.*

Мы проведём доказательство для случая матриц третьего порядка, но читатель может заметить, что дословно то же рассуждение можно провести и в случае любого порядка матриц-сомножителей.

Пусть даны матрицы

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \text{и} \quad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Их произведение будет:

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}, & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}, & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31}, & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32}, & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31}, & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32}, & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{pmatrix}. \tag{2}$$

Определитель этой последней матрицы *мы будем рассматривать как функцию столбцов матрицы* B. Обозначив указанные столбцы (которые мы будем рассматривать как трёхмерные числовые векторы) через $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$, можем записать определитель матрицы (2) как

$$F(\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3). \tag{3}$$

Обратим теперь внимание на свойства функции (3). Прежде всего очевидно, что если умножить один из столбцов матрицы B на

некоторое число, то на это же число умножится и соответствующий столбец матрицы (2), а следовательно, на то же число умножится значение определителя последней матрицы. Таким образом, например,

$$F(k\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)=kF(\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3).$$

Далее, если один из столбцов матрицы B, например первый, будет суммой двух столбцов $\boldsymbol{b}_1'$ и $\boldsymbol{b}_1''$, то соответствующий столбец матрицы (2) будет суммой столбца

$$a_{11}b_{11}'+a_{12}b_{21}'+a_{13}b_{31}',$$
$$a_{21}b_{11}'+a_{22}b_{21}'+a_{23}b_{31}',$$
$$a_{31}b_{11}'+a_{32}b_{21}'+a_{33}b_{31}'$$

и такого же столбца, содержащего b'' вместо b'. Пользуясь свойствами определителей, мы можем написать поэтому, что

$$F(\boldsymbol{b}_1'+\boldsymbol{b}_1'',\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)=F(\boldsymbol{b}_1',\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)+F(\boldsymbol{b}_1'',\ \boldsymbol{b}_2,\ \boldsymbol{b}_3).$$

Наконец, если два из столбцов матрицы B окажутся совпадающими, то будут совпадающими и соответствующие столбцы матрицы (2), а следовательно, определитель этой матрицы, т. е. значение $F(\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)$, обратится в нуль. Но эти свойства суть не что иное, как свойства А) и Б), использованные при введении понятия определителя. Поэтому к функции $F(\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)$ применима теорема, доказанная в конце § 4. Эта теорема даёт равенство

$$F(\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3)=\begin{vmatrix} b_{11} & b_{12} & b_{13}\\ b_{21} & b_{22} & b_{23}\\ b_{31} & b_{32} & b_{33}\end{vmatrix}F(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3),\qquad (3')$$

где $F(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3)$ есть значение нашей функции, когда столбцы матрицы B обращаются, соответственно, в

$$\boldsymbol{e}_1=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad \boldsymbol{e}_2=\begin{pmatrix}0\\1\\0\end{pmatrix}\quad\text{и}\quad \boldsymbol{e}_3=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

Но в этом случае матрица (2) совпадает с матрицей A, а следовательно, её определитель $F(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3)$ обращается в определитель матрицы A, т. е.

$$F(\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \boldsymbol{e}_3)=\begin{vmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix}.$$

Подставляя это в выражение (3'), получим равенство

$$F(\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3) = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \cdot \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

выражающее доказываемую теорему.

В заключение этого параграфа введём понятие об обратных матрицах. Мы уже заметили выше, что квадратные матрицы вида

$$E = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

при умножении на них не изменяют умножаемую матрицу. Матрица E называется единичной матрицей (конечно, для каждого порядка имеется своя единичная матрица). По аналогии со случаем чисел *матрица B называется обратной по отношению к матрице A, если оба произведения AB и BA равны единичной матрице.*

Отнюдь не все матрицы имеют обратные: из доказанной только что теоремы следует сразу, что *матрица A, определитель которой равен нулю* (*вырожденная* матрица), *не может иметь обратной*; в самом деле, если матрица имеет обратную матрицу B, то произведение определителей обеих матриц должно равняться определителю единичной матрицы, т. е. числу 1, а такое равенство невозможно, если хотя бы один из определителей матриц A и B равен нулю. Однако *любая матрица*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

определитель которой отличен от нуля, имеет обратную матрицу. Для того чтобы убедиться в этом, достаточно написать явное выражение для обратной матрицы. Именно, каждый элемент a_{ik} данной матрицы имеет в её определителе своё алгебраическое дополнение A_{ik}. Так как определитель матрицы A отличен от нуля, все эти алгебраические дополнения можно разделить на значение d определителя всей матрицы. Расставим эти частные таким образом:

$$B = \begin{pmatrix} \frac{A_{11}}{d} & \frac{A_{21}}{d} & \dots & \frac{A_{n1}}{d} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{d} & \frac{A_{2n}}{d} & \dots & \frac{A_{nn}}{d} \end{pmatrix} \tag{4}$$

(обращаем внимание читателя на то, что $\frac{A_{ik}}{d}$ стоит на пересечении k-й строки и i-го столбца новой матрицы). При образовании произведения матриц A и B можно заметить следующее: на пересечении i-й строки и j-го столбца произведения AB будет стоять сумма

$$\frac{a_{i1}A_{j1}+a_{i2}A_{j2}+\ldots+a_{in}A_{jn}}{d},$$

т. е. делённая на d сумма произведений элементов i-й строки матрицы A на алгебраические дополнения соответствующих элементов j-й строки той же матрицы. В силу известного свойства определителей такая сумма произведений равна нулю, если номера строк различны, и равна определителю, если номера строк совпадают. В последнем случае деление на d даёт единицу, так что в произведении рассматриваемых матриц на диагонали оказываются повсюду единицы, а все элементы, стоящие вне диагонали, — будут нулями. Иными словами, при умножении мы получаем единичную матрицу. Предоставляем читателю проверить, что единичная матрица получается и при умножении в другом порядке.

Во всём дальнейшем обратная матрица для матрицы A будет обозначаться (там, где она существует) через A^{-1}.

В наших новых обозначениях приобретают особую простоту свойства ортогональных матриц, рассмотренные в предыдущем параграфе. В рассмотренных там случаях плоскости и трёхмерного пространства ортогональные матрицы оказались такими, у которых суммы произведений соответствующих элементов различных столбцов равны нулю, а сумма квадратов элементов любого столбца равна единице. Если C есть такая матрица и если рассмотреть её транспонированную матрицу C^T, то мы легко убедимся, что произведение C^TC равно единичной матрице: умножать элементы i-й строки транспонированной матрицы на соответствующие элементы j-го столбца матрицы C означает то же самое, что и перемножать соответствующие элементы i-го и j-го столбцов матрицы C. Так как единицы получаются при этом в случае одинаковых номеров, а нули — в случае различных, то мы получаем указанный результат: $C^TC=E$.

Понятие ортогональной матрицы переносится на случай квадратных матриц любого порядка: такая матрица называется ортогональной, если её обратная матрица совпадает с транспонированной: $C^T=C^{-1}$.

Выше было обнаружено в случае плоскости, что определитель ортогональной матрицы может иметь только значения $+1$ или -1.

Это — свойство всех ортогональных матриц: если матрица C ортогональна, то из равенства $C^TC=E$ следует, в силу только что доказанной теоремы об определителе произведения матриц, равенство $|C^T|\cdot|C|=1$, где $|C^T|$ и $|C|$ суть определители матриц C^T и C.

Но по одному из основных свойств определителя $|C^T| = |C|$, а следовательно, $|C|^2 = 1$, т. е. $|C| = \pm 1$.

Добавим к сказанному, что определение умножения матриц можно использовать более широко, чем это было сделано выше: можно, например, составлять матрицу из векторов и умножать её на числовую матрицу. Правила действия, имеющие место для матриц, при этом полностью сохраняются. Этим обстоятельством мы дальше будем пользоваться. Например, формулы (1) предыдущего параграфа, связывающие базисы e_1', e_2', e_3' и e_1, e_2, e_3 пространства, запишутся в этом обозначении так:

$$(e_1', e_2', e_3') = (e_1, e_2, e_3) \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}. \qquad (5)$$

Нужно только помнить, что умножение двух матриц, составленных из векторов, не имеет смысла, а также то, что сложение может быть осуществлено только в случае однородных по своему характеру слагаемых.

§ 22. Линейные преобразования

Идея преобразования или отображения является одной из руководящих идей не только геометрии, но и всей математики. Её зарождение можно проследить уже на первых рисунках первобытного человека, в которых, несмотря на их примитивность, можно отчётливо увидеть стремление сопоставить каждой детали изображаемого предмета некоторый её «образ» на рисунке. Эта же идея многократно используется в школьном курсе геометрии, когда для доказательства ряда теорем и при решении задач используется движение фигур или переход от одной фигуры к другой, ей подобной.

Определение отображения в его наиболее общей и в то же время наиболее отчётливой форме может быть сформулировано так: пусть M и N — множества, состоящие из предметов (элементов) совершенно произвольной природы. Будем говорить, что нам задано некоторое *отображение множества M в множество N*, если указано правило, относящее каждому элементу множества M некоторый вполне определённый элемент множества N, называемый *образом* рассматриваемого элемента первого множества.

Для обозначения отображений часто применяется то же обозначение, что и для функций в математическом анализе: говорят, например, что задано отображение F, а образ элемента x множества M при этом отображении обозначают $F(x)$. Такое совпадение обозначений не случайно, так как обычные функции являются просто частными примерами отображений: каждая такая функция есть отображение некоторого множества чисел, на котором

она определена, на некоторое другое множество чисел. Например, функция $y=\arcsin x$ ставит в соответствие каждому числу, содержащемуся между -1 и $+1$, некоторое вполне определённое число, содержащееся между $-\frac{\pi}{2}$ и $\frac{\pi}{2}$. Когда мы в элементарной геометрии говорим о движении той или иной фигуры, то при этом имеем в виду, что заданная нам первоначальная фигура переведена в некоторую другую фигуру, причём эта новая фигура обладает тем свойством, что каждая точка на ней определённым образом соответствует некоторой точке исходной фигуры. Хотя в школьном курсе и не говорится явно о таком соответствии, его наличие на самом деле используется буквально во всех доказательствах, в которых применяется движение (достаточно вспомнить хотя бы доказательства признаков равенства треугольников, проводимые с помощью совмещения фигур).

Для наших целей необходимо в двух отношениях отклониться от элементарного представления о соответствии, используемого в школьном курсе: 1) когда мы будем говорить об отображении, то будем предполагать, что *оно определено во всём рассматриваемом пространстве,* т. е. что каждый элемент пространства (а не только элементы какой-либо фигуры) имеет определённый образ, и 2) в качестве элементов, между которыми устанавливается соответствие, берутся не точки, а векторы.

Таким образом, мы будем говорить, что задано отображение A векторного пространства L_1 в векторное пространство L_2, если каждому вектору x первого пространства поставлен в соответствие определённый вектор второго пространства, обозначаемый $A(x)$ или просто Ax.

В соответствии с применённой выше терминологией вектор Ax будет называться *образом* вектора x. Мы будем говорить также, что отображение A переводит вектор x в вектор Ax. Так как при заданном отображении A каждая система векторов преобразуется в некоторую другую систему, то отображения называются также *преобразованиями* рассматриваемого пространства.

Среди отображений векторных пространств особенно простыми и наиболее часто встречающимися в различных приложениях являются так называемые *линейные* отображения (или преобразования).

Отображение A векторного пространства L_1 в векторное пространство L_2 называется линейным, если оно обладает следующими двумя свойствами: 1) *образом суммы двух любых векторов является сумма их образов* и 2) *образом произведения вектора на любое число является произведение образа этого вектора на то же самое число.* На языке формул эти свойства записываются таким образом:

$$A(x+y)=Ax+Ay, \tag{1}$$

$$A(kx)=kAx. \tag{2}$$

Приведём несколько примеров линейных преобразований:

Пример 1. Поставим в соответствие каждому вектору трёхмерного пространства (в смысле обычной элементарной геометрии) его проекцию на одну из плоскостей (рис. 9). Этим, очевидно,

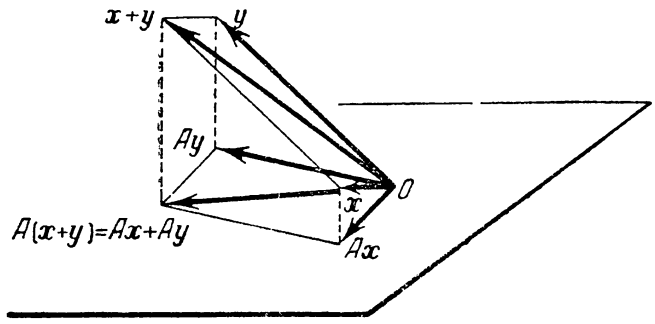

Рис. 9.

определяется отображение всего пространства на двумерное пространство L_2 векторов рассматриваемой плоскости. То, что определённое так отображение является линейным, следует из хорошо известных теорем о пропорциональных отрезках и из того, что проекции параллельных отрезков параллельны между собой.

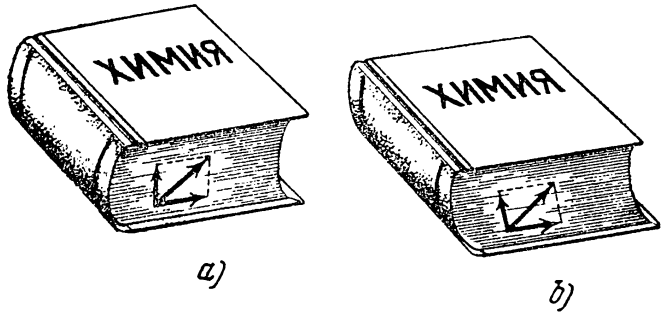

Рис. 10.

Пример 2. Отнесём каждому вектору плоскости другой вектор, получаемый из первого поворотом на постоянный угол φ (в одну и ту же сторону для всех векторов). Линейность получаемого таким образом отображения очевидна сразу.

Пример 3. Будем представлять себе плоскость реализованной в виде бокового среза книги (рис. 10), неподвижно лежащей на столе. Если мы слегка нажмём справа налево на корешок книги, то происходит сдвиг отдельных листов книги друг относительно друга. Такой сдвиг осуществляет линейное преобразование на нашей плоскости, как видно из сравнения рис. 10, *а* и *б*. Наш пример, строго

говоря, является лишь грубой иллюстрацией того, что следует называть сдвигом, так как из-за наличия отдельных листов книги наша «плоскость» приобретает дискретный характер. Однако получающееся наглядное представление весьма точно соответствует тому, что понимается под сдвигом в строгом геометрическом смысле слова.

Пример 4. Будем считать, что каждому вектору плоскости отнесено его «зеркальное отражение» в некоторой фиксированной прямой, т. е. вектор, расположенный симметрично с данным относительно этой прямой. Этим опять-таки определяется линейное отображение плоскости на ту же самую плоскость (рис. 11).

В трёх последних примерах имело место следующее частное обстоятельство: рассматриваемое векторное пространство отображалось в себя самого, т. е. исходные векторы и их образы оказались принадлежащими одному и тому же пространству. Этот частный случай с алгебраической точки зрения является наиболее интересным (и общий случай к нему легко сводится). Поэтому во всём дальнейшем, говоря о линейных преобразованиях, мы будем иметь в виду только линейные отображения пространства в себя.

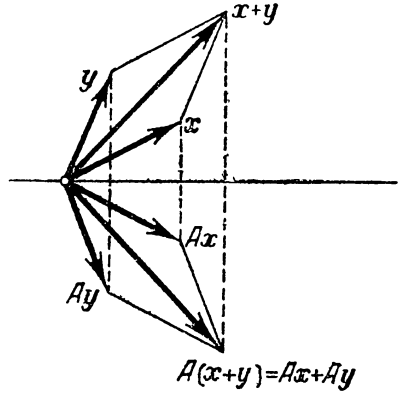

Рис. 11.

Связь введённых геометрических понятий с алгеброй устанавливается тем, что из заданных линейных преобразований можно создавать новые линейные преобразования с помощью формальных операций (имеющих, впрочем, вполне конкретное геометрическое содержание). После введения операций над преобразованиями последние становятся естественным объектом изучения алгебры.

Рассмотрим сначала сложение линейных преобразований. Пусть A и B — два линейных отображения некоторого векторного пространства L в себя. Эти преобразования ставят в соответствие каждому вектору x нашего пространства его образы Ax и Bx. Так как любые векторы пространства мы можем сложить, то можно образовать вектор $Ax+Bx$ и считать его соответствующим вектору x. Этим устанавливается некоторое новое отображение нашего пространства в себя, так как вектор $Ax+Bx$ определён для каждого вектора x. Обозначим полученное отображение через C. Оно будет линейным отображением. Для доказательства достаточно проверить наличие у него двух основных свойств (1) и (2) таких отображений. Но эти свойства

очевидны почти сразу: по определению отображения C мы имеем для любых векторов x и y такие соотношения:

$$C(x+y)=A(x+y)+B(x+y)=Ax+Ay+Bx+By,$$
$$C(kx)=A(kx)+B(kx)=kAx+kBx=k(Ax+Bx),$$
$$C(x)=Ax+Bx,\quad Cy=Ay+By,$$

откуда видно, что имеют место равенства

$$C(x+y)=Cx+Cy,\quad C(kx)=kCx,$$

содержание которых совпадает с содержанием равенств (1) и (2).

Определённое указанным образом отображение C называется суммой отображений A и B. Для обозначения суммы отображений применяется обычный знак: $C=A+B$. Знак равенства применяется здесь опять в смысле совпадения стоящих по обе стороны объектов (два преобразования естественно считать совпадающими, если они переводят каждый вектор пространства в один и тот же вектор).

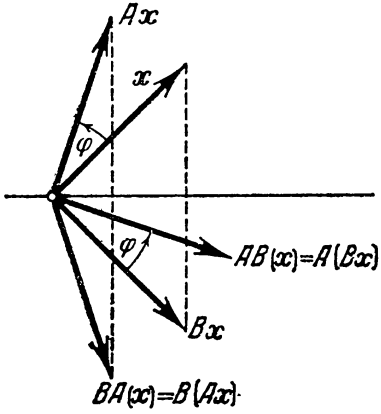

Рис. 12.

Умножение отображений вводится следующим образом: если A и B — два данных отображения, то, взяв некоторый вектор x пространства, можно перевести его сначала в вектор Bx, а затем к полученному вектору Bx применить преобразование A. Получим вектор $A(Bx)$. Описанный процесс ставит в соответствие каждому вектору x вектор $A(Bx)$ того же пространства, т. е. определяет некоторое отображение рассматриваемого пространства в себя. Это отображение и называется *произведением заданных отображений A и B*. Оно обозначается AB, причём *важно обращать внимание на порядок множителей*. Существенность этого замечания показывается хотя бы таким примером: пусть отображение A является вращением плоскости на угол φ (против часовой стрелки), а отображение B — зеркальным отражением в оси Ox (рис. 12). Читатель без труда убедится, что вектор x, если к нему сначала применить вращение A, а затем отражение B, обратится в вектор $B(Ax)$. Наоборот, если сначала к вектору x применить отражение B, а затем вращение A, то мы получим вектор $A(Bx)$, отличный от $B(Ax)$.

Произведение преобразований часто определяют короче и более выразительно, но менее точно, как *преобразование, эквивалентное последовательному выполнению заданных преобразований*. Нужно

подчеркнуть только (выше это было сказано явно), что *сначала выполняется преобразование, стоящее в произведении вторым множителем.* Произведение линейных отображений также является линейным отображением.

Как и в случае матриц, отмеченная выше некоммутативность умножения является единственным отступлением правил действий с преобразованиями от обычных алгебраических правил. Все остальные из основных законов алгебры полностью сохраняются: для любых преобразований A, B, C имеют место формулы

$$A(B+C)=AB+AC,$$
$$(A+B)C=AC+BC,$$
$$(AB)C=A(BC).$$

Эти соотношения могут быть выведены из соответствующих соотношений для матриц с помощью приёма, которым мы займёмся в следующем параграфе. Однако для закона ассоциативности умножения полезно здесь привести прямое доказательство. Именно, каков бы ни был вектор $\boldsymbol{x}$, имеет место равенство

$$(AB)C\boldsymbol{x}=(AB)(C\boldsymbol{x})=A(B(C\boldsymbol{x})).$$

(Здесь мы воспользовались только определением умножения преобразований: чтобы получить результат действия преобразования $(AB)C$ на вектор $\boldsymbol{x}$, мы применяем к нему сначала преобразование C, а затем преобразование AB; далее, для получения результата действия на вектор $C\boldsymbol{x}$ преобразования AB сначала применяем к вектору $C\boldsymbol{x}$ преобразование B, а затем — преобразование A.) Пользуясь этими же соображениями, для вектора $A(BC)\boldsymbol{x}$ получаем такое выражение:

$$A(BC)\boldsymbol{x}=A((BC)\boldsymbol{x})=A(B(C\boldsymbol{x})),$$

которое, как сразу видно, совпадает с приведёнными выше. Таким образом, результат действия преобразований $(AB)C$ и $A(BC)$ на любой вектор $\boldsymbol{x}$ будет одним и тем же, т. е. эти преобразования совпадают.

§ 23. Представление линейных преобразований матрицами

Хотя в предыдущем параграфе и были определены операции над линейными преобразованиями, их выполнение в конкретных случаях не столь легко, как выполнение обычных арифметических действий над числами. Чтобы получить способ просто выполнять эти операции, необходимо свести их к действиям над числами, а для этого, прежде всего, необходимо иметь способ аналитической записи самих линейных преобразований.

Ключ к получению такой записи даёт следующее замечание: *линейное преобразование вполне определено, если известно, в какие векторы оно переводит векторы базиса пространства.* В самом деле, если известны образы $Ae_1, \ldots, Ae_n$ векторов базиса $e_1, \ldots, e_n$ пространства L, то для любого вектора $x = e_1x_1 + \ldots + e_nx_n$ этого пространства его образ будет вполне определён: в силу основных свойств линейных преобразований должны иметь место соотношения

$$Ax = A(e_1x_1 + \ldots + e_nx_n) = Ae_1 \cdot x_1 + \ldots + Ae_n \cdot x_n.$$

Рассмотрим для определённости случай пространства трёх измерений. В этом случае векторов базиса всего три: e_1, e_2, e_3. Их образы Ae_1, Ae_2, Ae_3 будут также векторами пространства, а следовательно, они представляются линейными комбинациями самих векторов базиса:

$$\left.\begin{aligned} Ae_1 &= e_1a_{11} + e_2a_{21} + e_3a_{31}, \\ Ae_2 &= e_1a_{12} + e_2a_{22} + e_3a_{32}, \\ Ae_3 &= e_1a_{13} + e_2a_{23} + e_3a_{33}. \end{aligned}\right\} \tag{1}$$

При этом для любого вектора x мы будем иметь:

$$Ax = Ae_1 \cdot x_1 + Ae_2 \cdot x_2 + Ae_3 \cdot x_3 = \sum_{i,\,k=1}^{3} e_ia_{ik}x_k. \tag{2}$$

Таким образом, преобразование A будет вполне определено, если известна матрица

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \tag{3}$$

составленная из коэффициентов a_{ik} формул (1). Она называется *матрицей линейного преобразования* A относительно базиса e_1, e_2, e_3.

Введённые в § 21 операции над матрицами позволяют ещё упростить нашу запись и сделать работу с ней совершенно автоматической. Для этого условимся *под результатом действия линейного преобразования на любую строку, составленную из векторов, понимать строку, составленную из образов заданных векторов,* т. е., выражая это формулой, условимся в том, что

$$A(e_1, e_2, e_3) = (Ae_1, Ae_2, Ae_3).$$

Тогда формулы (1) запишутся так:

$$A(e_1, e_2, e_3) = (e_1, e_2, e_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \tag{1'}$$

а если ещё ввести сокращение, обозначая матрицу (3) через M_A

(здесь A указывает то преобразование, матрица которого обозначена M_A), то получим ещё более простую запись

$$A(e_1, e_2, e_3) = (e_1, e_2, e_3) M_A. \qquad (1'')$$

Преимущества введённой записи обнаружатся ещё ясней, когда применим её к формуле (2). Она может быть теперь переписана в виде

$$Ax = A\left((e_1, e_2, e_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = (e_1, e_2, e_3) M_A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

А если обозначить столбец, составленный из координат вектора x через X, то в виде

$$Ax = A((e_1, e_2, e_3) X) = (e_1, e_2, e_3) M_A X. \qquad (4)$$

Последнее соотношение остаётся справедливым также и в том случае, когда X означает любую матрицу: в этом случае, если через X_1, X_2, X_3 обозначить отдельные столбцы этой матрицы, произведение строки (e_1, e_2, e_3) на X означает строку, составленную из отдельных произведений

$$(e_1, e_2, e_3) X_1, \quad (e_1, e_2, e_3) X_2, \quad (e_1, e_2, e_3) X_3.$$

Действие линейного преобразования A на такую строку сводится к его действию на каждый из векторов строки, что приводит к системе векторов

$$(e_1, e_2, e_3) M_A X_1, \quad (e_1, e_2, e_3) M_A X_2, \quad (e_1, e_2, e_3) M_A X_3,$$

которая в свою очередь может быть записана в виде

$$(e_1, e_2, e_3) M_A X.$$

Заметим теперь, что матрица M_A, входящая в формулу (1''), однозначно определена линейным преобразованием, так как её столбцы являются коэффициентами линейных комбинаций векторов базиса, выражающих образы этих векторов; коэффициенты же в выражении любого вектора через базис определены однозначно.

Это замечание позволяет совершенно автоматически получить выражения для матриц суммы и произведения двух данных линейных преобразований. Действительно, по определению суммы и произведения преобразований, пользуясь введёнными в § 21 действиями с матрицами, получим формулы

$$(A+B)(e_1, e_2, e_3) = A(e_1, e_2, e_3) + B(e_1, e_2, e_3) = \\ = (e_1, e_2, e_3) M_A + (e_1, e_2, e_3) M_B = (e_1, e_2, e_3)(M_A + M_B) \qquad (5)$$

и

$$(AB)(e_1, e_2, e_3) = A(B(e_1, e_2, e_3)) = \\ = A((e_1, e_2, e_3) M_B) = (e_1, e_2, e_3) M_A M_B. \qquad (6)$$

Но матрицы преобразований $A+B$ и AB суть такие матрицы,

которые удовлетворяют соотношениям

$$(A+B)(e_1,\ e_2,\ e_3)=(e_1,\ e_2,\ e_3)\,M_{A+B},$$
$$(AB)(e_1,\ e_2,\ e_3)=(e_1,\ e_2,\ e_3)\,M_{AB}.$$

Сравнивая эти соотношения с формулами (5) и (6), получим:

$$M_{A+B}=M_A+M_B \quad \text{и} \quad M_{AB}=M_AM_B.$$

Таким образом, получается основное правило: *матрицы суммы и произведения линейных преобразований равны, соответственно, сумме и произведению матриц данных преобразований.*

Этим все действия с линейными преобразованиями сведены к соответствующим действиям с матрицами, т. е. в конечном счёте с числами.

Мы рассмотрели случай трёхмерного пространства. Но на самом деле, как это легко видеть из изложенного, *все полученные результаты могут быть приложены к пространству любого числа измерений*: изменится только число строк и столбцов рассматриваемых матриц.

В качестве иллюстрации рассмотрим выражение некоторых линейных преобразований плоскости, на которой в качестве базиса выбраны два взаимно перпендикулярных вектора e_1, e_2 длины 1.

Рассмотрим сначала вращение плоскости на угол φ (против часовой стрелки). Если обозначить это вращение через A, то легко видеть, что образами Ae_1 и Ae_2 векторов e_1 и e_2 будут векторы e_1' и e_2' (см. рис. 8, *а*). Принимая во внимание выражение этих векторов через базис e_1, e_2, получим соотношение

$$A(e_1,\ e_2)=(e_1,\ e_2)\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix},$$

показывающее, что матрицей преобразования A будет матрица

$$M_A=\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}.$$

Другим линейным преобразованиям будут соответствовать другие матрицы. Например, отражению B плоскости в прямой, на которой лежит вектор e_1, будет соответствовать матрица

$$M_B=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},$$

ибо

$$B(e_1,\ e_2)=(e_1,-e_2)=(e_1,\ e_2)\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Наши соображения дают возможность ещё раз убедиться в том, что произведения AB и BA преобразований A и B различны: этим

произведениям соответствуют матрицы

$$\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}=\begin{pmatrix}\cos\varphi & \sin\varphi\\ \sin\varphi & -\cos\varphi\end{pmatrix}$$

и

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}=\begin{pmatrix}\cos\varphi & -\sin\varphi\\ -\sin\varphi & -\cos\varphi\end{pmatrix},$$

так что преобразования AB и BA не могут совпадать.

Рассмотрим ещё один пример: возьмём то же отражение B плоскости в прямой, содержащей вектор e_1 (см. рис. 8, *а*), но в качестве базиса плоскости возьмём векторы e'_1, e'_2, изображённые на этом рисунке. Эти векторы преобразованием B переводятся в векторы Be'_1, Be'_2, изображённые на рис. 8, *а* пунктиром. Их выражения через векторы e'_1 и e'_2 будут:

$$Be'_1=e'_1\cos 2\varphi-e'_2\sin 2\varphi,\quad Be'_2=-e'_1\sin 2\varphi-e'_2\cos 2\varphi.$$

Отсюда видно, что матрица преобразования B в нашем новом базисе будет:

$$M'_B=\begin{pmatrix}\cos 2\varphi & -\sin 2\varphi\\ -\sin 2\varphi & -\cos 2\varphi\end{pmatrix}.$$

Она отличается от матрицы того же преобразования, полученной относительно базиса e_1, e_2. Таким образом, *одно и то же линейное преобразование может выражаться различными матрицами в зависимости от выбора базиса пространства.*

Однако между матрицами, выражающими линейное преобразование B в различных базисах, существует простая связь. Её вывод не представляет трудности, и мы проведём его в общем случае, независимо от размерности пространства. Пусть $e_1, e_2, \ldots, e_n$ и $e'_1, \ldots, e'_n$ — два различных базиса пространства L и A — некоторое линейное преобразование этого пространства. Тогда матрицы, выражающие это линейное преобразование относительно данных базисов, определяются из соотношений

$$\left.\begin{aligned}A(e_1,\ldots,e_n)&=(e_1,\ldots,e_n)\,M_A,\\ A(e'_1,\ldots,e'_n)&=(e'_1,\ldots,e'_n)\,M'_A.\end{aligned}\right\}\qquad(7)$$

Вспомним теперь, что переход от базиса $e_1, \ldots, e_n$ к базису $e'_1, \ldots, e'_n$ характеризуется матрицей перехода C:

или

$$\left.\begin{aligned}(e'_1,\ldots,e'_n)&=(e_1,\ldots,e_n)\,C\\ (e_1,\ldots,e_n)&=(e'_1,\ldots,e'_n)\,C^{-1}.\end{aligned}\right\}\qquad(8)$$

Применяя преобразование A к первой из этих формул, мы получим соотношение

$$A(e'_1,\ldots,e'_n)=A((e_1,\ldots,e_n)\,C)=(e_1,\ldots,e_n)\,M_A\,C$$

(последнее равенство написано на основании формулы (4), в которой роль матрицы X играет матрица C). Подставляя теперь выражение $(e_1, \ldots, e_n)$, даваемое второй из формул (8), получим соотношение

$$A(e'_1, \ldots, e'_n) = (e'_1, \ldots, e'_n) C^{-1} M_A C.$$

Сравнение его со второй формулой (7) даёт равенство

$$M'_A = C^{-1} M_A C, \tag{9}$$

устанавливающее связь между матрицами M'_A и M_A.

Отметим в заключение этого параграфа ещё следующее: как мы видели, при заданном базисе пространства каждому линейному преобразованию соответствует определённая матрица; остаётся вопрос, каждая ли матрица соответствует некоторому линейному преобразованию? Оказывается, что это действительно так. Чтобы убедиться в сказанном, достаточно определить отображение пространства в себя таким образом: если M_A — любая матрица, то примем формулы (1) в качестве определений образов векторов $e_1, \ldots, e_n$ при нашем преобразовании. После этого образ любого вектора x определим с помощью формулы (2).

Определённое таким образом преобразование пространства будет линейным, так как координаты суммы двух векторов равны суммам координат слагаемых, а координаты произведения вектора на число равны произведениям координат данного вектора на это число. Матрица нашего преобразования, по его определению, совпадает с матрицей M_A.

§ 24. Геометрические свойства линейных преобразований и свойства представляющих их матриц

После того как каждому линейному преобразованию поставлена в соответствие некоторая матрица, естественно посмотреть, какая связь существует между свойствами преобразования и свойствами представляющей его матрицы.

Линейное преобразование пространства L в себя называется *вырожденным,* если оно обращает пространство в некоторую его часть. Например, преобразование, состоящее в проектировании всех векторов трёхмерного пространства на некоторую плоскость, является вырожденным, так как при этом совокупность всех векторов пространства превращается в совокупность векторов только одной плоскости. Наоборот, *преобразование,* которое переводит пространство *во всё пространство,* называется *невырожденным.* Примерами невырожденных преобразований являются такие преобразования, как вращение трёхмерного пространства или плоскости, сдвиг (см. пример 3 § 22) и многие другие.

Вырожденность или невырожденность линейного преобразования может быть очень легко обнаружена по его матрице:

Преобразование A тогда и только тогда является невырожденным, когда соответствующая ему матрица имеет определитель, не равный нулю.

В самом деле, если преобразование A вырожденное, то векторы $Ae_1, \ldots, Ae_n$ не могут быть линейно независимыми, так как в этом случае любой вектор y пространства можно было бы представить их линейной комбинацией $Ae_1x_1 + \ldots + Ae_nx_n$, а это означает, что вектор y является образом вектора $x = e_1x_1 + \ldots + e_nx_n$, так что совокупность образов векторов пространства заполняет всё пространство, вопреки предположению. В силу доказанного в главе II условия линейной независимости векторов в таком случае определитель матрицы, составленный из коэффициентов выражений векторов $Ae_1, \ldots, Ae_n$ через базис $e_1, \ldots, e_n$, должен быть равен нулю. Но этот определитель и является определителем матрицы преобразования A.

Наоборот, если определитель матрицы M_A равен нулю, то в силу того же условия линейной зависимости, векторы $Ae_1, \ldots, Ae_n$ будут линейно зависимыми, а следовательно, их линейные комбинации *не могут заполнять всё рассматриваемое пространство.* Если теперь заметить, что образ любого вектора $x = e_1x_1 + \ldots + e_nx_n$ является такой линейной комбинацией, то станет ясным, что это и означает вырожденность преобразования A.

Невырожденные преобразования обладают ещё одним свойством, которое, подобно только что указанному, является для них характеристическим:

Преобразование A тогда и только тогда является невырожденным, когда единственным вектором, обращающимся после преобразования в нулевой вектор, является сам нулевой вектор.

В самом деле, если векторы $Ae_1, \ldots, Ae_n$ линейно зависимы, то найдутся некоторые числа $x_1, \ldots, x_n$, не все равные нулю, для которых $Ae_1 \cdot x_1 + \ldots + Ae_n \cdot x_n = 0$. Но это означает, что образ отличного от нуля вектора $x = e_1x_1 + \ldots + e_nx_n$ равен нулевому вектору. Наоборот, если векторы $Ae_1, \ldots, Ae_n$ линейно независимы, то образ $A(x) = Ae_1x_1 + \ldots + Ae_nx_n$ любого вектора $x = e_1x_1 + \ldots + e_nx_n$, отличного от нуля, сам не равен нулю.

Простейшим из всех преобразований является *тождественное* преобразование E, т. е. такое, которое оставляет все векторы пространства неподвижными: $Ex = x$. Из соотношений

$$E(e_1, \ldots, e_n) = (e_1, \ldots, e_n) = (e_1, \ldots, e_n) \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

следует, что *тождественное преобразование во всяком базисе выражается единичной матрицей.*

В случае невырожденного преобразования A (и только в этом случае) можно говорить об обратном преобразовании: это — такое преобразование, которое переводит образ Ax любого вектора x, полученный при преобразовании A, снова в вектор x. Другими словами, *обратным преобразованием для преобразования A называется такое преобразование A^{-1}, для которого произведение $A^{-1}A$ является тождественным преобразованием.* Установленное раньше правило образования матрицы произведения преобразований показывает, что матрицей обратного преобразования является обратная матрица матрицы данного преобразования, так как должно быть:

$$M_{A^{-1}}M_A = M_{A^{-1}A} = M_E = E.$$

Ограничимся снова рассмотрением преобразований плоскости и обычного трёхмерного пространства. В этих случаях, как мы отмечали выше, естественно пользоваться ортонормальными базисами (т. е. базисами, состоящими из взаимно перпендикулярных единичных векторов).

При пользовании такими базисами в совокупности всех линейных преобразований естественно выделить некоторые частные их типы. Формальным основанием такого выделения могут послужить свойства матриц, представляющих эти преобразования.

Важнейшими из определяемых таким путём классов преобразований являются следующие:

1. *Ортогональные преобразования,* т. е. линейные преобразования рассматриваемого пространства, *которые в некотором ортонормальном базисе выражаются ортогональными матрицами.*

2. *Симметрические преобразования, представляемые в таком базисе симметричной матрицей* (т. е. такой, которая не меняется при транспонировании: $M_A^T = M_A$).

Первый вопрос, который возникает в связи с таким определением: может ли преобразование представляться симметричной или ортонормальной матрицей в одном ортонормальном базисе и в то же время в другом базисе выражаться матрицей какого-либо другого характера. Ответ на этот вопрос — отрицательный:

Если некоторое линейное преобразование A представляется в ортонормальном базисе симметричной (или ортогональной) матрицей, то такой же матрицей оно представляется и в любом другом ортонормальном базисе.

В самом деле, переход от одного ортонормального базиса к другому даётся ортонормальной матрицей перехода C (см. § 21). Поэтому $C^{-1} = C^T$. Пользуясь теперь правилом, связывающим матрицы M_A и M'_A преобразования A в двух рассматриваемых базисах, будем иметь:

$$M'_A = C^{-1}M_AC = C^TM_AC.$$

Если матрица M_A была симметричной, т. е. если $M_A^T = M_A$, то матрица M'_A будет симметричной, ибо

$$M_A'^T = (C^T M_A C)^T = C^T M_A^T C^{TT} = C^T M_A C$$

(здесь мы пользовались образованием транспонированной матрицы для произведения матриц, указанным в § 21, а также тем обстоятельством, что транспонированная матрица для транспонированной матрицы совпадает с самой данной матрицей).

Если же матрица M_A была ортогональной, то из равенства $M_A^T M_A = E$, являющегося определением ортогональной матрицы, и из ортогональности матрицы C вытекают равенства

$$M_A'^T M'_A = (C^{-1} M_A C)^T (C^{-1} M_A C) = (C^T M_A C)^T (C^T M_A C) =$$
$$= C^T M_A^T C\, C^T M_A C = C^T M_A^T M_A C = C^T C = E,$$

т. е. ортогональность матрицы M'_A.

Ортогональные преобразования могут быть охарактеризованы геометрически:

Теорема. *Преобразование A тогда и только тогда является ортогональным, когда оно не меняет длин векторов, т. е. когда для любого вектора $\boldsymbol{x}$*

$$(A\boldsymbol{x},\ A\boldsymbol{x}) = (\boldsymbol{x},\ \boldsymbol{x}).$$

В самом деле, если преобразование A является ортогональным, то для любых векторов $\boldsymbol{e}_i$ и $\boldsymbol{e}_j$ ортонормального базиса имеют место равенства

$$(A\boldsymbol{e}_i,\ A\boldsymbol{e}_i) = 1 \quad \text{и} \quad (A\boldsymbol{e}_i,\ A\boldsymbol{e}_j) = 0 \quad \text{при} \quad i \neq j$$

(ведь именно эти равенства и выражаются условиями ортогональности матрицы). Отсюда следует, что для любого вектора

$$\boldsymbol{x} = \boldsymbol{e}_1 x_1 + \ldots + \boldsymbol{e}_n x_n$$

будет:

$$(A\boldsymbol{x},\ A\boldsymbol{x}) = x_1^2 + \ldots + x_n^2 = (\boldsymbol{x},\ \boldsymbol{x}).$$

Доказательство обратного утверждения несколько сложнее.

Заметим, прежде всего, что преобразование, не изменяющее скалярных произведений любых векторов, будет ортогональным: если $(A\boldsymbol{x},\ A\boldsymbol{y}) = (\boldsymbol{x},\ \boldsymbol{y})$ для любых векторов $\boldsymbol{x}$ и $\boldsymbol{y}$, то, в частности, $(A\boldsymbol{e}_i,\ A\boldsymbol{e}_j) = (\boldsymbol{e}_i,\ \boldsymbol{e}_j)$, т. е. будет равно нулю или единице, смотря по тому, будет ли $i \neq j$ или $i = j$. Но это и означает ортогональность матрицы преобразования.

Для полного доказательства формулированной теоремы остаётся доказать лишь такое утверждение:

Если линейное преобразование A не меняет длин векторов, то оно не меняет и их скалярных произведений.

В самом деле, если преобразование A не меняет длин векторов, то для любых векторов x и y должны выполняться равенства

$$(A(x+y),\ A(x+y))=(x+y,\ x+y),$$
$$(Ax,\ Ax)=(x,\ x),\quad (Ay,\ Ay)=(y,\ y).$$

Пользуясь теперь тем, что

$$(A(x+y),\ A(x+y))=(Ax+Ay,\ Ax+Ay)=$$
$$=(Ax,\ Ax)+2(Ax,\ Ay)+(Ay,\ Ay)$$

и

$$(x+y,\ x+y)=(x,\ x)+2(x,\ y)+(y,\ y),$$

из этих равенств получим $(Ax,\ Ay)=(x,\ y)$, т. е. скалярное произведение любых векторов равно скалярному произведению их образов. А это и нужно было доказать.

Легко убедиться теперь в том, что любое ортогональное преобразование пространства сводится к вращению и, может быть, ещё зеркальному отражению.

Ограничимся рассмотрением трёхмерного пространства.

Если e_1, e_2, e_3 — исходный ортонормальный базис и e_1', e_2', e_3' — образы его векторов, то очевидно следующее:

1) Можно произвести вращение пространства около начала координат так, чтобы вектор e_1 перешёл в e_1' (длины этих векторов равны). При этом векторы e_2 и e_3 также повернутся и станут перпендикулярными к e_1'.

2) После этого можно вращать пространство вокруг вектора e_1' так, чтобы вектор e_2 перешёл в e_2'.

Так как при этом вектор e_3 перейдёт в вектор, перпендикулярный к векторам e_1' и e_2', то возможны только два случая: либо повёрнутый таким образом вектор совпадёт с e_3', либо окажется противоположным ему. В последнем случае для достижения совпадения нужно произвести зеркальное отражение в плоскости, натянутой на два первых вектора. Таким образом, всегда можно достигнуть того, чтобы в результате вращения и некоторого зеркального отражения (если последнее нужно) векторы e_1, e_2, e_3 перешли в e_1', e_2' и e_3'. Так как вращение и зеркальное отражение являются линейными преобразованиями, то результирующее преобразование будет линейным. А так как его действие на векторы базиса e_1, e_2, e_3 таково же, что и в случае заданного преобразования A, то это результирующее преобразование будет совпадать с A.

§ 25. Симметрические преобразования. Случай плоскости

Геометрическая характеристика симметрических преобразований получается несколько сложнее, хотя результаты здесь столь же наглядны. Начнём со случая линейных преобразований плоскости, так как нужные нам вычисления здесь легко проводятся.

Пусть дано симметрическое преобразование A плоскости. Выбирая в качестве базиса два взаимно перпендикулярных единичных вектора, e_1, e_2, выразим наше линейное преобразование матрицей

$$A(e_1, e_2) = (e_1, e_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

При этом в силу симметричности преобразования $a_{12} = a_{21}$. Поставим вопрос: *не существует ли на плоскости такого вектора* $x = e_1x_1 + e_2x_2$, *который при преобразовании не меняет своего направления?* Предполагается, что этот вектор отличен от нуля, так как в противном случае не имеет смысла говорить о его направлении. Поставленное условие означает, что вектор Ax должен получаться из вектора x умножением на некоторое число λ. Записывая равенство $Ax = \lambda x$ в матричной форме по нашим правилам, получим:

$$A\left((e_1, e_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = (e_1, e_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (e_1, e_2)\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}.$$

Отсюда следует, что должны соблюдаться такие числовые равенства:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 &= \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 &= \lambda x_2, \end{aligned}\right\} \quad \text{или} \quad \left.\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 &= 0, \\ a_{12}x_1 + (a_{22} - \lambda)x_2 &= 0. \end{aligned}\right\} \tag{1}$$

Последние представляют собою систему линейных однородных уравнений с неизвестными x_1, x_2. Для того чтобы такая система имела отличное от нуля решение, как мы знаем, должно выполняться равенство определителя системы нулю

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} = 0.$$

Это даёт уравнение для определения возможных значений числа λ:

$$\lambda^2 - (a_{11} + a_{22})\lambda - a_{12}^2 + a_{11}a_{22} = 0. \tag{2}$$

Обращая теперь внимание на то, что коэффициенты этого уравнения действительны (мы рассматриваем обычную плоскость элементарной геометрии, так что все координаты векторов выражаются действительными числами) и что дискриминант уравнения

$$(a_{11} + a_{22})^2 + 4a_{12}^2 - 4a_{11}a_{22} = (a_{11} - a_{22})^2 + 4a_{12}^2 \geqslant 0,$$

можем утверждать, что корни уравнения (2) всегда действительны. Рассмотрим теперь отдельно два случая:

а) Корни уравнения (2) равны между собой. Это происходит тогда, когда дискриминант уравнения равен нулю, т. е. когда $(a_{11} - a_{22})^2 + 4a_{12}^2 = 0$. Но последнее равенство означает, что

$a_{11}-a_{22}=0$ и $a_{12}=0$. Таким образом, наше преобразование в рассматриваемом случае задаётся матрицей

$$\begin{pmatrix} a_{11} & 0 \\ 0 & a_{11} \end{pmatrix}.$$

Оно сводится к тому, что оба вектора базиса умножаются на одно и то же число a_{11}. Но в таком случае, как легко видеть, для любого вектора x будет: $Ax=a_{11}x$. Выражаясь геометрически, это означает, что A есть *преобразование подобия* с коэффициентами подобия a_{11}.

б) Корни уравнения (2) различны. Пусть они равны, соответственно, λ_1 и λ_2. Выражая коэффициенты уравнения через его корни, будем иметь:

$$\begin{aligned} a_{11}+a_{22}&=\lambda_1+\lambda_2, \\ a_{11}a_{22}-a_{12}^2&=\lambda_1\lambda_2. \end{aligned} \tag{3}$$

Подставим в уравнения (1) сначала одно из значений λ, например λ_1. Так как определитель системы уравнений (1) при этом обращается в нуль, система должна сводиться к одному уравнению, например к первому. Решение системы поэтому будет даваться значениями $x_1=a_{12}$ и $x_2=\lambda_1-a_{11}$ (или любыми значениями, им пропорциональными). Таким образом, мы находим один вектор

$$x_1=e_1a_{12}+e_2(\lambda_1-a_{11}),$$

который под действием преобразования A только умножается на число λ_1. Этим свойством будет обладать также любой другой вектор, совпадающий по направлению с вектором x_1. Аналогичным образом находим и другой вектор x_2, обладающий тем свойством, что $Ax_2=\lambda_2x_2$. Таким вектором будет, например, $x_2=e_1a_{12}+e_2(\lambda_2-a_{11})$. Легко усмотреть, что найденные векторы взаимно перпендикулярны, так как их скалярное произведение

$$\begin{aligned} (x_1, x_2)&=a_{12}^2+(\lambda_1-a_{11})(\lambda_2-a_{11})= \\ &=a_{12}^2+\lambda_1\lambda_2-(\lambda_1+\lambda_2)a_{11}+a_{11}^2=0\ ^{1}). \end{aligned}$$

Чтобы резюмировать полученные результаты, введём одно новое понятие: *вектор x, обладающий тем свойством, что $Ax=\lambda x$, называется собственным вектором преобразования A*. Число λ называется *собственным значением преобразования A*, которому принадлежит собственный вектор x.

Если использовать введённое понятие, то просмотр полученных выше результатов убеждает нас в справедливости следующей теоремы:

Теорема. *Каково бы ни было симметрическое преобразование плоскости, для него существуют два взаимно перпендикулярных*

[1]) Последнее получается применением формул (3).

собственных вектора, принадлежащих одному и тому же или различным собственным значениям.

Учитывая теперь возможность произвольно изменить длину собственного вектора, можем составить ортонормальный базис плоскости из собственных векторов заданного симметрического преобразования. Если обозначить векторы этого базиса через e_1' и e_2', а собственные значения, к которым они принадлежат, — через λ_1 и λ_2, то для рассматриваемого линейного преобразования получим такое матричное выражение:

$$A(e_1',\ e_2') = (\lambda_1 e_1',\ \lambda_2 e_2') = (e_1',\ e_2') \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Написанное выражение показывает, что *симметрическое преобразование сводится всегда к растяжению* (или сжатию) *плоскости в двух взаимно перпендикулярных направлениях.* Правда, это должно пониматься в несколько расширенном смысле, так как значения λ_1 и λ_2 могут быть отрицательными, а следовательно, направление одного или обоих векторов e_1' и e_2' при нашем преобразовании может измениться на противоположное. Растяжение или сжатие в обычном смысле получается в том случае, когда оба собственных значения λ_1 и λ_2 положительны.

§ 26. Симметрические преобразования трёхмерного пространства

Обратимся к случаю трёхмерного пространства. Здесь возможен тот же путь исследования, что и выше, но из-за большего числа координат векторов возникают некоторые осложнения. Чтобы запись нужных нам формул не была слишком громоздкой, приходится вводить некоторые вспомогательные средства формального характера.

Прежде всего введём понятие сопряжённого преобразования: если в некотором ортонормальном базисе задано преобразование A, матрица которого есть M_A, то преобразованием, *сопряжённым к* A, называется преобразование, матрицей которого будет M_A^T, т. е. транспонированная матрица преобразования A. Преобразование, сопряжённое преобразованию A, обозначается через A^*.

Это определение не зависит от выбранного ортонормального базиса: если перейти от заданного базиса к другому с помощью ортогональной матрицы перехода C, то матрица исходного преобразования A обратится в $C^T M_A C$, а матрица преобразования A^* — в матрицу $C^T M_A^T C$ ($C^{-1} = C^T$ и матрица перехода — одна и та же для всех преобразований). Но последняя матрица есть не что иное, как транспонированная новая матрица преобразования A: $(C^T M_A C)^T = C^T M_A^T C$. Таким образом, каким бы ортонормальным базисом мы ни воспользовались, мы получаем одно и то же преобразование A^*.

Сопряжённое преобразование A^* может быть также охарактеризовано следующим способом:

Преобразование A^* является сопряжённым с A в том и только в том случае, когда $(Ae_i, e_j)=(e_i, Ae_j^*)$ для любых векторов ортонормального базиса e_1, e_2, e_3.

В самом деле, пусть преобразования A и A^* заданы матричными соотношениями

$$A(e_1, e_2, e_3)=(e_1, e_2, e_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix};$$

$$A(e_1, e_2, e_3)=(e_1, e_2, e_3)\begin{pmatrix} a_{11}^* & a_{12}^* & a_{13}^* \\ a_{21}^* & a_{22}^* & a_{23}^* \\ a_{31}^* & a_{32}^* & a_{33}^* \end{pmatrix}.$$

Из этих соотношений видно, что

$$Ae_i=e_1a_{1i}+e_2a_{2i}+e_3a_{3i}$$

и

$$A^*e_j=e_1a_{1j}^*+e_2a_{2j}^*+e_3a_{3j}^* \qquad (i, j=1, 2, 3),$$

а следовательно, $(Ae_i, e_j)=a_{ji}$ и $(e_i, A^*e_j)=a_{ij}^*$. Отсюда видно, что указанное в формулировке доказываемого предложения равенство эквивалентно равенству $a_{ji}=a_{ij}^*$, т. е. условию, чтобы матрица преобразования A была транспонированной матрицей преобразования A. Эта характеристика может быть ещё несколько изменена и освобождена от какой бы то ни было зависимости от базиса пространства:

Преобразование A^* тогда и только тогда является сопряжённым к преобразованию A, когда для любых векторов имеет место равенство $(Ax, y)=(x, Ay^*)$.

В самом деле, если это равенство имеет место для любых векторов, то оно, в частности, справедливо для векторов базиса e_1, e_2, e_3, и мы приходим к доказанному предложению. Наоборот, если имеют место все равенства $(Ae_i, e_j)=(e_i, A^*e_j)$, то для любых векторов $x=e_1x_1+e_2x_2+e_3x_3$ и $y=e_1y_1+e_2y_2+e_3y_3$ в силу основных свойств линейных преобразований и свойств скалярного произведения будут выполняться равенства

$$(Ax, y)=(A(e_1x_1+e_2x_2+e_3x_3), e_1y_1+e_2y_2+e_3y_3)=$$
$$=\sum_{i, j=1}^{3} x_iy_j(Ae_i, e_j),$$

$$(x, A^*y)=(e_1x_1+e_2x_2+e_3x_3, A^*(e_1y_1+e_2y_2+e_3y_3))=$$
$$=\sum_{i, j=1}^{3} x_iy_j(e_i, A^*e_j),$$

Отсюда следует, что если имеют место все равенства $(Ae_i, e_j) = (e_i, A^*e_j)$, то должно быть:

$$(Ax, y) = (x, A^*y).$$

Связь введённых определений с симметрическими преобразованиями очевидна: само определение симметрического преобразования показывает, что оно эквивалентно такому:

Линейное преобразование A называется симметрическим, если оно совпадает со своим сопряжённым преобразованием: $A = A^*$.

Ортогональные преобразования могут быть также определены с помощью понятия сопряжённого преобразования:

Преобразование A будет ортогональным тогда и только тогда, когда его обратное преобразование совпадает с сопряжённым, т. е. когда $A^*A = E$.

В самом деле, в этом случае для матриц преобразований A и A^* будет справедливо соотношение $M_{A^*}M_A = M_A^T M_A = E$, т. е. матрица M_A будет ортогональной.

Всё сказанное относится также и к случаю плоскости. Однако в этом случае нет особой необходимости во введённых понятиях, так как имеющие здесь место соотношения достаточно просты, и их можно исследовать непосредственно, как это было сделано выше.

Рассмотрим теперь произвольное симметрическое преобразование трёхмерного пространства, заданное в некотором ортонормальном базисе матричными соотношениями

$$A(e_1, e_2, e_3) = (e_1, e_2, e_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \qquad (a_{ik} = a_{ki}).$$

Сохраняя прежнее определение собственного вектора и собственного значения, можем получить в координатной форме условия того, чтобы вектор $x = e_1x_1 + e_2x_2 + e_3x_3$ был собственным вектором преобразования A, принадлежащим собственному значению λ. Эти условия будут иметь вид

$$A\left((e_1, e_2, e_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = (e_1, e_2, e_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} =$$

$$= (e_1, e_2, e_3)\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix}$$

и приводятся к системе уравнений

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= \lambda x_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= \lambda x_3, \end{aligned}\right\}$$

т. е.

$$\left.\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + a_{13}x_3 &= 0, \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + a_{23}x_3 &= 0, \\ a_{31}x_1 + a_{32}x_2 + (a_{33} - \lambda)x_3 &= 0. \end{aligned}\right\} \qquad (1)$$

Для того чтобы последняя система имела решение, отличное от нулевого, необходимо и достаточно обращение определителя её в нуль:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0. \qquad (2)$$

Это — уравнение третьей степени. Но известно, что уравнение третьей степени с действительными коэффициентами имеет хотя бы один действительный корень (многочлен третьей степени принимает значения разных знаков при больших положительных и больших по абсолютной величине отрицательных значениях неизвестного). Таким образом, найдётся хотя бы одно значение λ, удовлетворяющее уравнению (2). При этом значении вторая из систем уравнений (2) допускает решение, отличное от нулевого, и значения x_1, x_2, x_3, удовлетворяющие этой системе, будут координатами некоторого собственного вектора $\boldsymbol{x}$ нашего преобразования. Таким образом, получаем следующую лемму:

Лемма. *Любое симметрическое линейное преобразование трёхмерного пространства имеет по крайней мере один собственный вектор.*

Это позволяет немедленно обнаружить существование *трёх взаимно перпендикулярных собственных векторов* следующим образом. Обозначим найденный собственный вектор (будем предполагать, что он сделан по длине равным единице) через $\boldsymbol{e}_1'$. Рассмотрим совокупность всех векторов, перпендикулярных к вектору $\boldsymbol{e}_1'$. Они заполняют некоторую плоскость, которую мы будем теперь рассматривать как самостоятельное *двумерное векторное пространство. Эта плоскость замечательна тем, что её векторы при применении к ним преобразования A остаются перпендикулярными к вектору* $\boldsymbol{e}_1'$, *т. е. остаются лежащими на нашей плоскости.* В самом деле, если $\boldsymbol{x}$ — вектор плоскости, т. е. если $(\boldsymbol{x}, \boldsymbol{e}_1') = 0$, то $(A\boldsymbol{x}, \boldsymbol{e}_1') = (\boldsymbol{x}, A^*\boldsymbol{e}_1') = (\boldsymbol{x}, A\boldsymbol{e}_1') = (\boldsymbol{x}, \lambda_1 \boldsymbol{e}_1') = 0$, так

как в силу симметрии преобразования $A^* = A$. Но равенство скалярного произведения (Ax, e_1') нулю означает, что вектор Ax перпендикулярен к вектору e_1' (может быть, впрочем, и $Ax = 0$, но это не меняет нашего общего вывода, так как нулевой вектор лежит на плоскости).

Полученный результат означает, что наше преобразование можно рассматривать также *как преобразование упомянутой плоскости в себя*. Оно будет линейным, так как оно линейно во всём пространстве. Кроме того, это преобразование плоскости будет симметрическим: равенство $(Ax, y) = (x, Ay)$, вытекающее из условия $A = A^*$, имеет место даже для всех векторов пространства, а не только для векторов плоскости. Но выше мы доказали, что для любого симметрического преобразования плоскости на ней найдутся два взаимно перпендикулярных собственных вектора e_2', e_3' (их также можно предполагать равными единице по длине).

Таким образом, в пространстве найдены три взаимно перпендикулярных собственных единичных вектора e_1', e_2', e_3' преобразования A. Их можно принять за базис пространства. В этом базисе преобразование A задаётся соотношением

$$A(e_1', e_2', e_3') = (e_1', e_2', e_3') \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix},$$

показывающим, что преобразование снова сводится по существу к простому растяжению пространства в трёх взаимно перпендикулярных направлениях (с возможным изменением этих направлений на противоположные).

Растяжения пространства в собственном смысле выделяются среди всех симметрических преобразований тем, что для них все собственные значения λ_1, λ_2, λ_3 положительны. Этого же можно достигнуть с помощью понятия о положительно определённом преобразовании: *преобразование* A называется *положительно определённым*, если ни один вектор не образует со своим образом тупого угла, т. е. если для любого вектора x имеет место соотношение $(Ax, x) \geqslant 0$.

Собственные значения положительно определённого симметрического преобразования не отрицательны.

В самом деле, если e_1 — собственный вектор преобразования A, то из неравенства $(Ae_1, e_1) = (\lambda_1 e_1, e_1) \geqslant 0$ и из того, что скалярное произведение (e_1, e_1) как квадрат длины вектора e_1 положительно, вытекает, что $\lambda_1 \geqslant 0$.

Таким образом, растяжения пространства характеризуются тем, что они являются положительно определёнными симметрическими преобразованиями.

§ 27. Представление произвольного линейного преобразования произведением ортогонального и симметрического

Из результатов предыдущего параграфа вытекает следующая теорема, показывающая, что произвольное линейное преобразование пространства (или плоскости) сводится к последовательному выполнению растяжения в трёх взаимно перпендикулярных направлениях, вращению и, может быть, отражению в некоторой плоскости.

Теорема. *Любое невырожденное линейное преобразование является произведением симметрического преобразования и некоторого ортогонального.*

Доказательство этой замечательной теоремы удобно разбить на несколько этапов.

1. *Произведение линейного преобразования на его сопряжённое является симметрическим преобразованием, причём положительно определённым.*

Если преобразованию A соответствует в некотором ортонормальном базисе матрица M_A, то преобразованию A^* в этом же базисе соответствует матрица M_A^T. Отсюда следует, что произведению A^*A соответствует матрица $M_A^T M_A$. Её транспонированная матрица $(M_A^T M_A)^T = M_A^T (M_A^T)^T$ совпадает с ней, так что преобразование A^*A симметрично.

Чтобы доказать положительную определённость преобразования A^*A, достаточно воспользоваться свойством сопряжённого преобразования: если $\boldsymbol{x}$ — произвольный вектор, то

$$(\boldsymbol{x},\ A^*A\boldsymbol{x}) = (A\boldsymbol{x},\ A\boldsymbol{x}).$$

Но скалярное произведение вектора на себя не может быть отрицательным, так что $(\boldsymbol{x},\ A^*A\boldsymbol{x}) \geqslant 0$.

2. *Для любого симметрического положительно определённого преобразования существует другое положительно определённое преобразование, квадрат которого равен данному.*

Согласно доказанному в предыдущем параграфе в пространстве найдётся такой ортонормальный базис, в котором матрица данного симметрического преобразования имеет вид

$$M_A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Числа λ_1, λ_2, λ_3 не могут быть отрицательными в силу положительной определённости данного преобразования. Поэтому можно

образовать матрицу

$$M_B = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{pmatrix},$$

которая, как и всякая матрица, будет матрицей некоторого линейного преобразования B. В силу того, что эта матрица симметрична, преобразование B будет симметрическим. Квадрат преобразования B будет совпадать с данным преобразованием, так как произведение матрицы B на себя даёт матрицу M_A.

3. *Сопряжённое преобразование для произведения двух или нескольких преобразований равно произведению их сопряжённых в обратном порядке.*

Это утверждение просто следует из рассмотрения матриц: если A и B — данные преобразования, а M_A и M_B — матрицы, соответствующие им в некотором ортонормальном базисе, то $M_A M_B$ будет матрицей преобразования AB, а $(M_A M_B)^T$ — матрицей сопряжённого преобразования $(AB)^*$. Но $(M_A M_B)^T = M_B^T M_A^T$, а это есть не что иное, как произведение матриц M_{B^*} и M_{A^*} преобразований B^* и A^*.

После всего сказанного доказательство формулированной теоремы не представляет труда. Пусть A — данное невырожденное преобразование. Образуем произведение A^*A. Это преобразование будет симметрическим, положительно определённым и невырожденным (последнее следует из теоремы об определителе произведения двух матриц). В силу только что доказанного, существует симметрическое положительно определённое преобразование B, квадрат которого равен A^*A:

$$B^2 = A^*A;$$

преобразование B будет также невырожденным. Поэтому существует обратное преобразование B^{-1}, а следовательно, мы можем написать равенство

$$A = (AB^{-1})\,B.$$

Формулированная теорема будет доказана, если мы обнаружим, что преобразование AB^{-1} будет ортогональным. Но мы имеем:

$$(AB^{-1})^* AB^{-1} = B^{-1}A^*AB^{-1} = B^{-1}B^2B^{-1} = E\ ^1),$$

откуда следует, что преобразование $(AB^{-1})^*$ является обратным для преобразования AB^{-1}, а это и нужно (см. замечание об ортогональных преобразованиях в предыдущем параграфе).

---

[1]) Преобразование B^{-1} будет также симметрическим: $(B^{-1})^* = B^{-1}$.

§ 28. Упрощение уравнений линий и поверхностей второго порядка.

Изложенная теория допускает важные приложения к исследованию линий и поверхностей второго порядка.

Мы начнём с рассмотрения кривых, причём ограничимся частным случаем, когда кривая имеет центр симметрии. Такие кривые называются *центральными*.

Если воспользоваться любой прямоугольной координатной системой, начало которой расположено в центре симметрии кривой второго порядка, то её уравнение запишется в виде

$$ax^2 + 2bxy + cy^2 = d, \tag{1}$$

где множитель 2 при произведении координат введён для удобства.

Обозначим через e_1 и e_2 единичные векторы, направленные вдоль координатных осей. Они образуют базис рассматриваемой плоскости. Если x и y — координаты некоторой точки, то $r = e_1x + e_2y$ — радиус-вектор.

Рассмотрим теперь линейное преобразование A плоскости, определяемое в рассматриваемом базисе матрицей.

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

составленной из коэффициентов уравнения кривой. Легко видеть, что для вектора $A(r)$ будем иметь выражение

$$A(r) = A(e_1)x + A(e_2)y =$$
$$= (ae_1 + be_2)x + (be_1 + ce_2)y = e_1(ax + by) + e_2(bx + cy).$$

Отсюда следует, что левая часть уравнения (1) является скалярным произведением векторов r и $A(r)$, так что уравнение рассматриваемой кривой может быть переписано в виде

$$(r, Ar) = d. \tag{2}$$

Такая запись позволяет легко судить о том, что происходит с уравнением кривой при переходе от одной координатной системы к другой. В силу того, что преобразование A является симметрическим, на плоскости существует ортонормальный базис, состоящий из собственных векторов e_1' и e_2' преобразования A. Если перейти к системе координат, определяемой этим базисом, то мы будем иметь соотношения

$$r = e_1'x' + e_2'y', \quad A(e_1') = \lambda_1 e_1',$$
$$A(e_2') = \lambda_2 e_2', \quad A(r) = e_1'\lambda_1 x' + e_2'\lambda_2 y'$$

и

$$(e_1', e_1') = (e_2', e_2') = 1, \quad (e_1', e_2') = 0.$$

Поэтому уравнение (2) перепишется в виде

$$\lambda_1 x'^2 + \lambda_2 y'^2 = d. \tag{3}$$

Это и будет уравнением нашей кривой в новой системе координат.

Простота формы уравнения (3) даёт возможность по нему непосредственно судить о форме кривой, что было бы затруднительным сделать, пользуясь первоначальным уравнением (1). В частности, из уравнения (3) сразу видно, что новые оси координат являются осями симметрии рассматриваемой кривой. Этим доказано существование у любой центральной кривой второго порядка двух взаимно перпендикулярных осей симметрии. Можно было бы, несколько дополнив приведённые соображения, получить полную классификацию всех кривых второго порядка. Однако это увело бы нас от основной темы.

Сила применённого метода станет ещё более наглядной при рассмотрении случая центральных поверхностей второго порядка.

Если начало координат находится в центре симметрии такой поверхности, то её уравнение будет иметь вид

$$ax^2 + 2bxy + cy^2 + 2dxz + 2fyz + gz^2 = h.$$

Читателю предоставляется самому проверить, что это уравнение может быть записано в виде $(\boldsymbol{r}, A(\boldsymbol{r})) = h$, если через $\boldsymbol{r}$ обозначить радиус-вектор точки, а через A — симметрическое преобразование, определяемое матрицей

$$\begin{pmatrix} a & b & d \\ b & c & f \\ d & f & g \end{pmatrix}.$$

Если теперь найти ортонормальный базис пространства, состоящий из собственных векторов преобразования A, и перейти к системе координат, определяемой этим базисом, то повторение приведённых выше рассуждений показывает сразу, что в новой координатной системе уравнение нашей поверхности будет иметь вид

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 = h.$$

Как и в случае кривых, полученное уравнение даёт возможность легко решать все вопросы, относящиеся к форме и расположению рассматриваемой поверхности. В частности, отсюда следует существование трёх взаимно перпендикулярных плоскостей симметрии у любой центральной поверхности второго порядка.

Литература

1. Шапиро Г. М., Высшая алгебра, изд. 4-е, дополненное, Учпедгиз, 1938.

2. Окунев Л. Я., Высшая алгебра, изд. 4-е, Гостехиздат, 1949.

3. Курош А. Г., Курс высшей алгебры, изд. 2-е, Гостехиздат, 1950.

4. Сушкевич А. К. Основы высшей алгебры, изд. 4-е, Гостехиздат, 1941.

В этих четырёх книгах можно найти изложение теории определителей, отличное от приведённого в настоящей статье.

Наиболее доступной по своему изложению является книга Г. М. Шапиро.

5. Виноградов С. П., Основания теории детерминантов, ГТТИ, 1933.

Кроме лёгкости изложения, эта книжечка обладает ещё тем достоинством, что в ней собрано довольно много поучительных задач, снабжённых ответами и иногда указаниями.

6. Каган В. Ф., Основы теории определителей, Одесса, 1922.

Обширная монография по вопросам, связанным с теорией определителей. В ней содержится, в частности, изложение многих исторически имевших место подходов к введению определителей.

7. Фаддеев Д. К. и Соминский И. С., Сборник задач по высшей алгебре, Гостехиздат, 1949.

В книге даны многочисленные примеры вычисления определителей. В нужных случаях задачи снабжены указаниями, так что книга может быть рекомендована всем, кто хочет овладеть техникой вычисления определителей.

8. Гельфанд И. М., Лекции по линейной алгебре, Гостехиздат, 1948.

9. Мальцев А. И., Линейная алгебра, Гостехиздат, 1948.

В этих двух книгах читатель найдёт более подробное изложение самой теории линейных преобразований, причём во второй из них изложение сопровождается некоторым количеством задач, разбор которых способствует усвоению освещаемого в книге материала.

Л. Я. ОКУНЕВ

КОЛЬЦО МНОГОЧЛЕНОВ И ПОЛЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ

*

ГЛАВА I

КОЛЬЦО МНОГОЧЛЕНОВ ОТ ОДНОГО НЕИЗВЕСТНОГО

§ 1. Кольцо многочленов

В курсе элементарной алгебры понятия многочлена и рациональной функции (алгебраической дроби) играют существенную роль. В средней школе часто приходится иметь дело с такими выражениями, как, например,

$$x^3+x^2,\quad x^2+xy+y^2+5,\quad \frac{x^2+y}{xy}$$

и т. п. Возникает естественный вопрос: что понимать под этими выражениями, в каком смысле следует понимать равенство, сложение и умножение подобного рода выражений?

Обратимся, прежде всего, к конкретным примерам; они позволят нам лучше уяснить те трудности, с которыми связано обоснование понятий многочлена и алгебраической дроби.

Возьмём многочлен x^3+x^2. Этот многочлен можно рассматривать как функцию переменного x, определённую на множестве действительных или на множестве комплексных чисел. Однако такая функциональная точка зрения, характерная для математического анализа, в алгебре в ряде случаев оказывается неприемлемой. Обратимся хотя бы к целым числам; обозначим через $\bar{0}$ совокупность чётных и через $\bar{1}$ — совокупность нечётных чисел и рассмотрим конечное множество P, состоящее только из двух элементов, а именно из $\bar{0}$ и $\bar{1}$. Установим во множестве P операции сложения и умножения с помощью равенств

$$\left.\begin{array}{lll} \bar{0}+\bar{0}=\bar{0}, & \bar{0}+\bar{1}=1+0=\bar{1}, & 1+\bar{1}=\bar{0}, \\ \bar{0}\cdot\bar{0}=\bar{0}, & \bar{0}\cdot\bar{1}=\bar{1}\cdot\bar{0}=\bar{0}, & \bar{1}\cdot\bar{1}=\bar{1}. \end{array}\right\} \qquad (1)$$

Определяя так операции сложения и умножения, мы руководствовались тем, что сумма двух чётных чисел есть число чётное, сумма чётного и нечётного есть число нечётное, сумма двух

нечётных чисел есть число чётное, произведение двух чётных — чётное и т. п.

Предоставляем читателю самому проверить, что относительно введённых операций сложения и умножения множество P образует поле[1]), причём $\overline{0}$ является нулём, а $\overline{1}$ — единицей поля.

Посмотрим теперь, что получится, если многочлен x^3+x^2 рассматривать как функцию переменного x, определённую на множестве P. Нетрудно заметить на основании равенств (1), что функция x^3+x^2 равна нулю на всём множестве P. Следовательно,

$$x^3+x^2=0.$$

Отсюда

$$x^3+x^2+x=x, \quad x^4+x^3+x^2=x^4$$

и т. п.

Итак, рассматривая многочлены как функции, определённые на множестве P, мы пришли к правилам обращения с многочленами, резко отличающимся от обычных.

В случае алгебраических дробей возникает ещё одно осложнение. Обратимся, например, к дробям

$$f(x)=\frac{x^2-1}{x-1}, \quad g(x)=\frac{(x+1)^2}{x+1}$$

и будем их рассматривать как функции в области действительных чисел. Тогда $f(x)$ и $g(x)$ будут иметь различные области определения: функция $f(x)$ определена для действительных значений x, отличных от 1, а функция $g(x)$ определена для действительных значений x, отличных от -1. Поэтому мы вынуждены рассматривать $f(x)$ и $g(x)$ как различные функции:

$$\frac{x^2-1}{x-1}\neq\frac{(x+1)^2}{x+1}.$$

Точно так же

$$\frac{x^2-1}{x-1}\neq x+1.$$

Таким образом, при изложении общей теории многочленов и алгебраических дробей целесообразнее отказаться от функциональной точки зрения, и мы соответствующие понятия будем вводить чисто алгебраически. Несколько ниже (в §§ 3, 11 и 12) мы установим, в каких случаях алгебраическая и функциональная точки зрения на многочлены и алгебраические дроби являются эквивалентными.

Своё изложение мы начнём с многочленов от одного x. Что касается многочленов от нескольких $x_1, \ldots, x_n$ и алгебраических

---

[1]) Определение кольца и поля, а также основные свойства кольца и поля см. Э. э. м., кн. 1, И. В. Проскуряков, Понятия множества, группы, кольца и поля. Теоретические основы арифметики.

дробей, то о них речь пойдёт лишь после изложения общей теории многочленов от одного x.

Понятие многочлена от x можно ввести следующим образом.

Пусть R — некоторое коммутативное кольцо с единицей $e \neq 0$. Элементы кольца R мы будем обозначать начальными буквами a, b, c, ... латинского алфавита. *Многочленом от x над кольцом R мы назовём выражение вида*

$$a_1x^{k_1} + a_2x^{k_2} + \ldots + a_sx^{k_s} \qquad (s \geqslant 1), \tag{2}$$

где a_1, a_2, ... , a_s — элементы из R, $k_1 < k_2 < \ldots < k_s$ — целые неотрицательные числа, x^0 принимается равным единице e, а также принимается, что при любом целом неотрицательном k $ex^k = x^ke = x^k$.

Следует обратить внимание на то обстоятельство, что мы здесь x, x^2, ... , x^k, ... , а также выражения $a_1x^{k_1}$, ... , $a_sx^{k_s}$ и знак $+$, соединяющий $a_1x^{k_1}$, ... , $a_sx^{k_s}$, рассматриваем как символы, которым не приписывается определённого значения. В соответствии с этим x будет называться *неизвестным*. В дальнейшем, после введения понятия равенства многочленов и действий сложения и умножения многочленов, мы придём ко вполне определённому истолкованию неизвестного x; символы x^k совпадут со степенями x, а само выражение (2) будет восприниматься как сумма произведений этих степеней на элементы из кольца R.

Отметим, что элементы кольца R можно всегда рассматривать как многочлены над R, а именно как многочлены вида ax^0. Очевидно, что выражение ax^k, где k — произвольное целое неотрицательное число, и, в частности, само неизвестное x являются также многочленами над R.

Элементы a_1, a_2, ... , a_s, входящие в выражение (2), обычно называются *коэффициентами*, а $a_1x^{k_1}$, $a_2x^{k_2}$, ... , $a_sx^{k_s}$ — *членами* многочлена (2). В частности, $a_sx^{k_s}$ называется *старшим членом* и a_s — старшим коэффициентом многочлена.

Для сокращения письма мы часто будем многочлены обозначать через $f(x)$, $g(x)$, $h(x)$ и т. п.

Введём понятия равенства, суммы и произведения многочленов от x над кольцом R.

Пусть $f(x)$ и $g(x)$ — два произвольных многочлена над R. Эти многочлены мы считаем равными (тождественно равными) только тогда, когда многочлен $f(x)$ состоит из тех же членов, что и многочлен $g(x)$, кроме членов с коэффициентами, равными нулю (если такие члены имеются). Например, многочлены

$$x + x^2 + x^3 \quad \text{и} \quad 0 + x + x^2 + x^3 + 0 \cdot x^4$$

равны. Напротив, многочлены

$$f(x) = x + x^2 + x^3 \quad \text{и} \quad g(x) = x + x^2 + x^3 + x^4$$

не равны, так как $g(x)$ обладает членом x^4, не входящим в состав членов $f(x)$.

Из этого определения равенства многочленов вытекает, что мы можем всякий многочлен $f(x)$ над R привести к виду

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n$$

(n — целое неотрицательное число),

добавляя, в случае необходимости, члены с коэффициентами, равными нулю. В таком виде мы часто будем записывать многочлен.

Согласно определению равенства многочленов имеем, в частности, что многочлен $f(x)$ равен нулю (т. е. нулевому элементу кольца R) лишь в том случае, когда все коэффициенты $f(x)$ равны нулю. Таким образом, если многочлен $f(x)$ не равен нулю, то по меньшей мере один из его коэффициентов должен быть отличен от нуля.

Обратимся теперь к действиям сложения и умножения многочленов. Пусть

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n,$$
$$g(x)=b_0+b_1x+b_2x^2+\ldots+b_mx^m$$

— два произвольных многочлена над R. Под их суммой $f(x)+g(x)$ мы будем подразумевать многочлен

$$d_0+d_1x+d_2x^2+\ldots+d_kx^k,$$

где k есть наибольшее из чисел n и m, $d_i=a_i+b_i$; при этом если $n>m$, то следует полагать: $b_{m+1}=\ldots=b_n=0$, а если $n<m$, то следует полагать: $a_{n+1}=\ldots=a_m=0$.

Под произведением $f(x)g(x)$ мы будем подразумевать многочлен

$$a_0b_0+(a_0b_1+a_1b_0)x+\ldots+(a_0b_k+a_1b_{k-1}+\ldots+a_kb_0)x^k+\ldots$$
$$\ldots+a_nb_mx^{n+m}, \quad \text{(A)}$$

где $a_i=0$ при $i>n$ и $b_j=0$ при $j>m$.

Посмотрим теперь, что вытекает из этих определений суммы и произведения многочленов над R.

Обозначим через $R[x]$ множество всех многочленов от x над кольцом R. Мы утверждаем, что введённые нами действия сложения и умножения подчиняются основным алгебраическим законам. Точнее, имеет место следующая

Теорема 1. *Множество $R[x]$ образует кольцо относительно сложения и умножения многочленов над R и притом кольцо коммутативное.*

Доказательство. Очевидно, что, складывая или перемножая два каких-нибудь многочлена от неизвестного x с коэффициентами из R, мы всегда получаем однозначно многочлен от x с коэффи-

циентами из того же кольца R. Таким образом, сложение и умножение многочленов от x над R являются алгебраическими операциями, определёнными во множестве $R[x]$.

Нетрудно, далее, проверить, что операции сложения и умножения многочленов из $R[x]$ подчиняются переместительному, сочетательному и распределительному законам. Мы ограничимся выводом сочетательности умножения.

Умножим

$$f(x)\,g(x) = a_0b_0 + (a_0b_1 + a_1b_0)\,x + \ldots + a_nb_mx^{n+m}$$

на многочлен

$$h(x) = c_0 + c_1x + \ldots + c_lx^l.$$

Согласно определению произведения многочленов получаем:

$$[f(x)\,g(x)]\,h(x) = \\ = a_0b_0c_0 + (a_0b_0c_1 + a_0b_1c_0 + a_1b_0c_0)\,x + \ldots + a_nb_mc_lx^{n+m+l}. \quad (3)$$

С другой стороны,

$$g(x)\,h(x) = b_0c_0 + (b_0c_1 + b_1c_0)\,x + \ldots + b_mc_lx^{m+l},$$

откуда на основании того же определения произведения

$$f(x)\,[g(x)\,h(x)] = \\ = a_0b_0c_0 + (a_0b_0c_1 + a_0b_1c_0 + a_1b_0c_0)\,x + \ldots + a_nb_mc_lx^{n+m+l}. \quad (4)$$

Сравнивая коэффициенты многочленов (3) и (4) при одинаковых x^i и вспоминая определение равенства многочленов, видим, что

$$[f(x)\,g(x)]\,h(x) = f(x)\,[g(x)\,h(x)],$$

т. е. умножение многочленов над R подчиняется сочетательному закону [1]).

Наконец, нетрудно убедиться, что во множестве $R[x]$ сложение всегда обратимо: для любых двух многочленов

$$f(x) = a_0 + a_1x + \ldots + a_nx^n$$

и

$$g(x) = b_0 + b_1x + \ldots + b_mx^m$$

из $R[x]$ уравнение $f(x) + z = g(x)$ всегда разрешимо в $R[x]$.

В самом деле, легко проверить, пользуясь определением сложения многочленов, что при $n = m$

$$z = (b_0 - a_0) + (b_1 - a_1)\,x + \ldots + (b_n - a_n)\,x^n,$$

---

[1]) Существенную роль в доказательстве теоремы играет то обстоятельство, что R есть кольцо. Например, при проверке сочетательного закона для умножения многочленов мы опирались на сочетательный и распределительный законы сложения и умножения элементов кольца R.

при $n > m$

$$z = (b_0 - a_0) + (b_1 - a_1)x + \ldots + (b_m - a_m)x^m + \\ + (-a_{m+1})x^{m+1} + \ldots + (-a_n)x^n$$

и при $n < m$

$$z = (b_0 - a_0) + (b_1 - a_1)x + \ldots + (b_n - a_n)x^n + \\ + (-b_{n+1})x^{n+1} + \ldots + (-b_m)x^m.$$

Только что доказанная теорема позволяет сделать ряд заключений относительно многочленов над R. Отметим наиболее существенные.

1° Благодаря выполнимости сочетательного закона для операции сложения мы можем теперь многочлен

$$f(x) = a_0 + a_1x + \ldots + a_nx^n$$

рассматривать как сумму его членов a_ix^i. При этом члены многочлена $f(x)$ можно записывать в любом порядке следования, так как сложение подчиняется и переместительному закону. Например, мы могли бы многочлен $f(x)$ записывать также и в порядке убывания индексов коэффициентов:

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \ldots + a_0.$$

2° Благодаря тому, что операция умножения подчиняется сочетательному закону, мы теперь можем символы x^2, x^3, ..., x^n рассматривать как степени неизвестного x, причём $x^sx^t = x^{s+t}$. Каждый член a_ix^i многочлена $f(x)$ можно рассматривать как произведение элемента a_i кольца R на степень x^i неизвестного x. При этом в силу переместительности умножения $a_ix^i = x^ia_i$.

3° Произведение $(ax^k)(bx^l)$, где a, b — элементы кольца R, k и l — целые неотрицательные числа, равно abx^{k+l}. Так как, сверх того, для операций сложения и умножения многочленов из $R[x]$ имеет место распределительный закон, то $f(x)$ и $g(x)$ можно перемножать по обычному школьному правилу, состоящему в том, что каждый член многочлена $f(x)$ умножается на каждый член многочлена $g(x)$, в результате чего получаются выражения вида $a_ib_jx^{i+j}$, составляется сумма всех таких выражений и, наконец, производится приведение подобных членов [1]).

4° Уравнение $f(x) + z = g(x)$, где $f(x)$ и $g(x)$ — произвольные многочлены из $R[x]$, имеет на основании известного свойства кольца единственное решение. Это единственное решение z будет

---

[1]) Законность приведения подобных членов основана на том, что операция сложения многочленов из $R[x]$ подчиняется переместительному и сочетательному законам и что сложение и умножение многочленов из $R[x]$ связаны распределительным законом.

обозначаться через $g(x)-f(x)$ и называться *разностью* многочленов $g(x)$ и $f(x)$. В частности, имеем

$$f(x)-f(x)=0, \quad 0-f(x)=-f(x),$$

где $-f(x)$—многочлен, противоположный $f(x)$, т. е. такой многочлен, сумма которого с многочленом $f(x)$ равна нулю.

Поскольку множество $R[x]$ образует кольцо относительно сложения и умножения многочленов над R, мы будем $R[x]$ называть *кольцом многочленов от x над R*.

До сих пор неизвестное x рассматривалось как чистый символ. Теперь мы можем дать неизвестному x некоторое истолкование. Для этой цели введём понятия подкольца, расширения кольца и трансцендентного элемента.

Пусть K — произвольное кольцо (коммутативное или некоммутативное — безразлично). Если некоторая часть K' элементов кольца K образует в свою очередь кольцо относительно тех же самых операций $+$ и $\cdot$, то мы эту часть назовём *подкольцом* кольца K, а само K — *расширением* кольца K'.

Например, кольцо чётных чисел является подкольцом кольца целых чисел, а кольцо целых чисел — расширением кольца чётных чисел. Другим примером может служить кольцо многочленов $R[x]$: оно является расширением кольца R.

Будем и в дальнейшем через R обозначать коммутативное кольцо с единицей $e\neq 0$. Пусть Ω — некоторое коммутативное расширение кольца R, обладающее той же самой единицей e, что и R. Элемент θ из Ω мы назовём *трансцендентным* относительно R, если для любого целого неотрицательного n равенство

$$a_0+a_1\theta+\ldots+a_n\theta^n=0,$$

где $a_0, a_1, \ldots, a_n$ — элементы из R, возможно лишь в том случае, когда $a_0=a_1=\ldots=a_n=0$.

Из этого определения видно, что трансцендентный элемент θ является по отношению к кольцу R внешним элементом: он не может лежать в R. В самом деле, если бы θ лежало в R, то $\theta=a$, где a — некоторый элемент кольца R, т. е. мы имели бы равенство $\theta-a=0$ с коэффициентом $e\neq 0$, что противоречит трансцендентности θ.

Исторически первым примером трансцендентных элементов были так называемые трансцендентные числа, т. е. комплексные числа, трансцендентные относительно кольца целых чисел. Впервые существование трансцендентных чисел было установлено Лиувиллем в 1851 г. В 1873 г. Эрмитом была обнаружена трансцендентность числа e, основания натуральных логарифмов. В 1882 г. Линдеман показал, что и число π, отношение длины окружности к диаметру,— также трансцендентно. Дальнейший существенный шаг вперёд в развитии теории трансцендентных чисел был сделан советским

учёным А. О. Гельфондом в 1929—1936 гг., установившим трансцендентность одного замечательного класса чисел [1]).

Возникает вопрос, существует ли для всякого кольца R трансцендентный элемент. Ответ на этот вопрос даёт следующая

Теорема 2 (о существовании трансцендентного элемента). *Для всякого коммутативного кольца R с единицей $e \neq 0$ существует коммутативное расширение Ω с той же самой единицей e, содержащее по меньшей мере один элемент, трансцендентный относительно R.*

Мы опускаем доказательство этой теоремы [2]).

Обозначим через $R[\theta]$ множество элементов расширения Ω кольца R, имеющих вид $a_0 + a_1\theta + \ldots + a_n\theta^n$ (n — произвольное целое неотрицательное число), где θ — попрежнему элемент из Ω, трансцендентный относительно R, а $a_0, a_1, \ldots, a_n$ — произвольные элементы из R.

Можно показать, что операции сложения и умножения элементов кольца Ω являются алгебраическими операциями и для $R[\theta]$.

В самом деле, пусть

$$\alpha = a_0 + a_1\theta + \ldots + a_n\theta^n,$$
$$\beta = b_0 + b_1\theta + \ldots + b_m\theta^m$$

и предположим для определённости, что $n \geqslant m$.

Обратимся к сумме $\alpha + \beta$:

$$\alpha + \beta = (a_0 + a_1\theta + \ldots + a_n\theta^n) + (b_0 + b_1\theta + \ldots + b_m\theta^m). \quad (5)$$

В Ω, как и во всяком кольце, сложение подчиняется сочетательному и переместительному законам. Следовательно, в правой части равенства (5) можно раскрыть скобки и сгруппировать слагаемые с одинаковыми степенями θ. В результате мы получим:

$$\alpha + \beta = (a_0 + b_0) + (a_1\theta + b_1\theta) + \ldots + (a_m\theta^m + b_m\theta^m) + a_{m+1}\theta^{m+1} + \ldots + a_n\theta^n \text{ [3])},$$

или на основании распределительного закона:

$$\alpha + \beta = (a_0 + b_0) + (a_1 + b_1)\theta + \ldots + (a_m + b_m)\theta^m + a_{m+1}\theta^{m+1} + \ldots + a_n\theta^n.$$

---

[1]) Более полные сведения о трансцендентных числах см. Э. э. м., кн. 1, А. Я. Хинчин, Элементы теории чисел.

[2]) Соответствующее доказательство см. в кн. И. В Проскурякова [6] на стр. 239—240, а также в книге Л. Я. Окунева [4] на стр. 330. Отметим попутно, что в теореме 2 говорится лишь о существовании расширения Ω, содержащего по меньшей мере один трансцендентный элемент, но ничего не говорится о природе такого расширения. Поэтому было бы ошибочным предполагать, что из теоремы 2 вытекает существование трансцендентных чисел.

[3]) Если $n = m$, то члены $a_{m+1}\theta^{m+1}, \ldots, a_n\theta^n$ отпадают.

Но a_0+b_0, a_1+b_1, ..., a_m+b_m, a_{m+1}, ..., a_n суть элементы R. Следовательно, $\alpha+\beta$ есть элемент множества $R[\theta]$.

Подобным же образом убеждаемся, что $\alpha\beta$ есть также элемент $R[\theta]$ (здесь, кроме переместительности и сочетательности сложения и распределительного закона, придётся воспользоваться сочетательностью и переместительностью умножения элементов кольца Ω).

Но ещё существеннее следующее обстоятельство.

Теорема 3. *Множество $R[\theta]$ и кольцо многочленов $R[x]$ изоморфны.*

Доказательство. Поставим в соответствие каждому многочлену $f(x)=a_0+a_1x+\ldots+a_nx^n$ из $R[x]$ элемент $\alpha=a_0+a_1\theta+\ldots+a_n\theta^n$ из $R[\theta]$ с теми же самыми a_0, a_1, ..., a_n, что и у многочлена $f(x)$:

$$f(x)=a_0+a_1x+\ldots+a_nx^n\rightarrow\alpha=a_0+a_1\theta+\ldots+a_n\theta^n. \quad (6)$$

($\rightarrow$ — знак соответствия).

Мы сейчас покажем, что соответствие (6) является изоморфизмом между $R[x]$ и $R[\theta]$.

Прежде всего нетрудно убедиться, что соответствие (6) не только однозначно, но и взаимно однозначно. Для этой цели возьмём из $R[x]$ ещё один многочлен

$$g(x)=b_0+b_1x+\ldots+b_mx^m.$$

Тогда

$$g(x)=b_0+b_1x+\ldots+b_mx^m\rightarrow\beta=b_0+b_1\theta+\ldots+b_m\theta^m.$$

Если $\alpha=\beta$, то $\alpha-\beta=0$. Но

$$\alpha-\beta=(a_0+a_1\theta+\ldots+a_n\theta^n)-(b_0+b_1\theta+\ldots+b_m\theta^m)=0.$$

Отсюда, опираясь на основные алгебраические свойства операций, имеющие место во всяком коммутативном кольце и, в частности, в кольце Ω, получаем, что при $n=m$

$$(a_0-b_0)+(a_1-b_1)\theta+\ldots+(a_n-b_n)\theta^n=0,$$

при $n>m$

$$(a_0-b_0)+(a_1-b_1)\theta+\ldots+(a_m-b_m)\theta^m+a_{m+1}\theta^{m+1}+\ldots+a_n\theta^n=0$$

и при $n<m$

$$(a_0-b_0)+(a_1-b_1)\theta+\ldots+(a_n-b_n)\theta^n-b_{n+1}\theta^{n+1}-\ldots-b_m\theta^m=0.$$

Так как θ трансцендентно относительно R, то получается, что при $n=m$

$$a_0=b_0,\quad a_1=b_1,\ \ldots,\ a_n=b_n,$$

при $n>m$

$$a_0=b_0,\quad a_1=b_1,\ \ldots,\ a_m=b_m,\ a_{m+1}=0,\ \ldots,\ a_n=0,$$

при $n<m$

$$a_0=b_0,\quad a_1=b_1,\ \ldots,\ a_n=b_n,\quad b_{n+1}=0,\ \ldots,\ b_m=0.$$

Мы видим отсюда, что $f(x)=g(x)$.

Для завершения доказательства теоремы остаётся показать, что сумме и произведению любых двух многочленов из $R[x]$ соответствуют сумма и произведение соответствующих элементов из $R[\theta]$.

Пусть, например,

$$f(x)=a_0+a_1x+\ldots+a_nx^n,$$
$$g(x)=b_0+b_1x+\ldots+b_mx^m$$

и $n\geqslant m$. Тогда согласно определению сложения многочленов из $R[x]$

$$f(x)+g(x)=c_0+c_1x+\ldots+c_nx^n,$$

где $c_i=a_i+b_i$, причём при $n>m$ коэффициенты $b_{m+1},\ \ldots,\ b_n$ следует считать равными нулю. Этому многочлену $f(x)+g(x)$ мы должны поставить в соответствие

$$\gamma=c_0+c_1\theta+\ldots+c_n\theta^n.$$

Но

$$\alpha+\beta=(a_0+a_1\theta+\ldots+a_n\theta^n)+(b_0+b_1\theta+\ldots+b_m\theta^m)=$$
$$=c_0+c_1\theta+\ldots+c_n\theta^n=\gamma.$$

Следовательно,

$$f(x)+g(x)\to\alpha+\beta.$$

Аналогичным образом обнаруживается, что

$$f(x)g(x)\to\alpha\beta.$$

Теорема доказана.

Из изоморфизма кольца многочленов $R[x]$ с множеством $R[\theta]$ вытекает, что $R[\theta]$ есть также кольцо. Тем самым $R[\theta]$ есть расширение кольца R, содержащее элемент θ, и в то же время $R[\theta]$ есть подкольцо кольца Ω. Оказывается, что $R[\theta]$ является минимальным в следующем смысле: *никакое подкольцо кольца $R[\theta]$, отличное от $R[\theta]$ и являющееся расширением R, уже не может содержать элемента θ.*

В самом деле, пусть S — некоторое подкольцо кольца $R[\theta]$, содержащее элемент θ и являющееся расширением R. Тогда, очевидно, S будет содержать не только θ, но и любую целую неотрицательную степень θ^k элемента θ. Так как S является расширением R, то S должно содержать произвольный элемент a кольца R, а потому должно содержать и $a\theta^k$. Отсюда S должно содержать и всевозможные элементы вида

$$a_0+a_1\theta+\ldots+a_n\theta^n,$$

где n — произвольное целое неотрицательное число, $a_0, a_1, \ldots, a_n$ — произвольные элементы из R. Таким образом, S должно совпадать с $R[\theta]$, и минимальность $R[\theta]$ становится очевидной.

Изоморфизм кольца многочленов $R[x]$ с кольцом $R[\theta]$ даёт нам право считать $R[x]$ и $R[\theta]$ неразличимыми с точки зрения их алгебраических свойств относительно операций сложения и умножения, а само x рассматривать как элемент, трансцендентный относительно R. Таково истолкование неизвестного x.

Введём теперь понятие степени многочлена от неизвестного x. Возьмём из $R[x]$ произвольный многочлен, не равный нулю. В таком многочлене по меньшей мере один коэффициент должен быть отличен от нуля. Назовём степенью этого многочлена наибольшую из степеней его членов, у которых коэффициенты не равны нулю.

Например,

$$f(x) = 1 + 2x + 3x^2 + 0 \cdot x^3 + 2x^4 + 0 \cdot x^5$$

есть многочлен от x четвёртой степени над кольцом целых чисел.

Если степень многочлена $f(x)$ равна n, то, очевидно, мы его можем всегда записать в виде

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n$$

или в виде

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_0$$

со старшим коэффициентом a_n, отличным от нуля, так как члены, содержащие x в степени выше n-й, равны нулю, и мы их можем отбросить.

Всякий элемент $a \neq 0$ кольца R можно рассматривать как многочлен нулевой степени от неизвестного x, потому что $a = ax^0$. Что касается нуля кольца R, то мы его будем рассматривать как многочлен, не имеющий степени.

Понятие степени позволяет весьма просто выразить условие равенства двух многочленов от x. А именно, если

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n \qquad (a_n \neq 0)$$

и

$$g(x) = b_0 + b_1 x + \ldots + b_m x^m \qquad (b_m \neq 0)$$

— два многочлена из $R[x]$ соответственно степени n и m, то эти многочлены равны лишь в том и только в том случае, когда их степени равны и равны их коэффициенты при одинаковых степенях неизвестного:

$$n = m, \quad a_0 = b_0, \quad a_1 = b_1, \quad \ldots, \quad a_n = b_n.$$

Что касается многочлена, равного нулю, то, как мы знаем, все коэффициенты такого многочлена должны быть равны нулю.

Из определения операции сложения многочленов от x легко усмотреть, что степень суммы $f(x) + g(x)$ не превосходит степени

каждого из слагаемых $f(x)$ и $g(x)$; она может оказаться и ниже степеней $f(x)$ и $g(x)$. Например, если

$$f(x)=2x^3+x^2+x-1, \quad g(x)=-2x^3-x^2+5x+6,$$

то

$$f(x)+g(x)=6x+5$$

будет уже многочленом первой степени.

С первого взгляда на равенство (А) (см. стр. 132) может показаться, что степень произведения многочленов $f(x)$ и $g(x)$ равна сумме их степеней. Но это заключение в случае произвольного кольца R ошибочно. Дело в том, что существуют кольца R с делителями нуля, т. е. кольца, в которых произведение элементов может равняться нулю и в том случае, когда сомножители отличны от нуля: $ab=0$ и при $a\neq 0$, $b\neq 0$. Такие элементы $a\neq 0$ и $b\neq 0$, как известно, называются *делителями нуля*.

Одним из простейших примеров кольца с делителями нуля может служить множество всех матриц второго порядка

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

с действительными элементами a_{11}, a_{12}, a_{21}, a_{22}. Нетрудно убедиться, что это множество образует кольцо относительно сложения и умножения матриц [1]). Вместе с тем легко видеть, что в этом кольце роль нуля играет нулевая матрица второго порядка, т. е. матрица

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

все элементы которой равны нулю. Возьмём теперь следующие две матрицы:

$$a=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad b=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Эти матрицы отличны от нуля, так как каждая из них содержит в качестве элемента число 1, не равное нулю. Однако их произведение согласно правилу перемножения матриц будет равно нулю, т. е. будет нулевой матрицей.

Таким образом, если кольцо R обладает делителями нуля и

$$f(x)=a_0+a_1x+\ldots+a_nx^n \qquad (a_n\neq 0)$$

— многочлен n-й степени над R, а

$$g(x)=b_0+b_1x+\ldots+b_mx^m \qquad (b_m\neq 0)$$

---

[1]) О действиях над матрицами см. в § 21 статьи А. И. Узкова «Векторные пространства и линейные преобразования».

— многочлен m-й степени над R, причём старшие коэффициенты a_n и b_m являются делителями нуля: $a_n b_m = 0$, то степень произведения $f(x)g(x)$ будет уже меньше $n+m$, так как член $a_n b_m x^{n+m}$ равен нулю. Но *если хотя бы один из старших коэффициентов a_n или b_m не является делителем нуля, то степень произведения $f(x)g(x)$ будет в точности равна сумме степеней $n+m$ сомножителей.*

В случае если R — область целостности (т. е. кольцо без делителей нуля) и, в частности, поле[1]), то степень произведения $f(x)g(x)$ многочленов $f(x)$ и $g(x)$ над R всегда равна сумме степеней сомножителей.

Отметим ещё одно существенное свойство многочленов от x над областью целостности R.

Теорема 4. *Если кольцо R является областью целостности, то кольцо многочленов $R[x]$ также является областью целостности.*

Доказательство. Пусть $f(x)$ и $g(x)$ — два многочлена из $R[x]$, не равные нулю. Эти многочлены должны иметь вполне определённую степень. Пусть степень $f(x)$ равна n, а степень $g(x)$ равна m. Тогда степень произведения $f(x)g(x)$ будет равна $n+m$, так как R по условию есть область целостности. Мы видим, что произведение $f(x)g(x)$ имеет вполне определённую степень, в силу чего $f(x)g(x) \neq 0$. Итак, в кольце $R[x]$ не существует делителей нуля; кольцо $R[x]$, следовательно, есть область целостности.

В заключение введём понятие значения многочлена от неизвестного x над R. Оно будет играть немаловажную роль.

Пусть

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n$$

— произвольный многочлен из $R[x]$. Заменим в нём неизвестное x каким-нибудь элементом c кольца R. Мы получим элемент того же кольца R следующего вида:

$$d = a_0 + a_1 c + \ldots + a_n c^n.$$

Этот элемент d называется *значением* многочлена $f(x)$ при значении неизвестного $x = c$ и обозначается через $f(c)$. Подчёркиваем, что под значением неизвестного x мы всегда будем подразумевать тот или иной элемент кольца R.

Очевидно, что если $f(x) = g(x)$, то $f(c) = g(c)$ для любого c из R. Обратное утверждение вообще неверно: мы знаем, что многочлены $f(x)$ и $g(x)$ над конечным кольцом могут быть и не равны, и тем не менее $f(c)$ будет равно $g(c)$ для любого c из R (см. хотя бы пример, приведённый на стр. 129). Впрочем, в § 3 мы увидим,

---

[1]) Что поле не содержит делителей нуля, можно убедиться следующим образом. Пусть $ab = 0$ и $a \neq 0$. Тогда $a^{-1}ab = 0$ или $b = 0$, так как $a^{-1}a = e$ — единица поля.

что для частного случая, когда R является бесконечной областью целостности [1]), обратное утверждение верно, и мы увидим, что для таких колец функциональная точка зрения на многочлен является оправданной. Это обстоятельство играет важную роль при изучении свойств многочленов над полем действительных и над полем комплексных чисел.

Нетрудно проверить, что если

$$f(x)+g(x)=h(x), \quad f(x)\,g(x)=k(x),$$

то

$$f(c)+g(c)=h(c), \quad f(c)\,g(c)=k(c). \tag{7}$$

Мы ограничимся проверкой первого равенства (7).

Пусть, например, степень многочлена $f(x)$ больше или равна степени многочлена $g(x)$:

$$f(x)=a_0+a_1x+\ldots+a_nx^n \qquad (a_n\neq 0),$$
$$g(x)=b_0+b_1x+\ldots+b_mx^m \qquad (b_m\neq 0)$$

и $m\leqslant n$. Тогда

$$h(x)=c_0+c_1x+\ldots+c_nx^n,$$

где $c_i=a_i+b_i$, причём в случае $n>m$ надо считать $b_{m+1},\ldots,b_n$ равными нулю. Найдём, чему равно $h(c)$:

$$h(c)=c_0+c_1c+\ldots+c_nc^n.$$

Пользуясь переместительным, сочетательным и распределительным законами, имеющими место в R, мы можем последнее равенство преобразовать в следующее:

$$h(c)=(a_0+a_1c+\ldots+a_nc^n)+(b_0+b_1c+\ldots+b_mc^m),$$

т. е. получилось, что

$$h(c)=f(c)+g(c).$$

Подобным же образом проверяется и второе соотношение (7).

§ 2. Свойства делимости многочленов от одного неизвестного

Большой интерес в алгебре представляет кольцо многочленов с коэффициентами из того или другого поля; в этом случае наблюдается далеко идущая аналогия между свойствами делимости многочленов и свойствами делимости целых чисел. В настоящем параграфе мы изложим в основных чертах теорию делимости

[1]) То-есть областью целостности, состоящей из бесконечного множества элементов.

многочленов над произвольным полем P. Единицу поля P мы будем обозначать через 1.

Ещё в определении поля под делением подразумевалось нахождение корня уравнения $ax=b$. Назовём в соответствии с этим многочлен $f(x)$ из кольца $P[x]$ делящимся на многочлен $g(x)\neq 0$ из того же кольца $P[x]$, если уравнение $g(x)X=f(x)$ разрешимо в $P[x]$. Иными словами, многочлен $f(x)$ делится на многочлен $g(x)$, если в том же кольце существует такой третий многочлен $q(x)$, что $f(x)=g(x)q(x)$.

Не следует думать, однако, что в кольце многочленов $P[x]$ деление всегда выполнимо. Возьмём хотя бы многочлены $f(x)=x+1$ и $g(x)=x^2+1$. Легко видеть, что $f(x)$ не делится на $g(x)$. В самом деле, если бы $f(x)$ делилось на $g(x)$, то в $P[x]$ нашёлся бы многочлен $q(x)$, для которого имело бы место равенство

$$x+1=(x^2+1)q(x).$$

Но это равенство невозможно, так как степень произведения $(x^2+1)q(x)$ выше степени $x+1$. Таким образом, $P[x]$, подобно множеству целых чисел, есть область целостности, но не поле. Отметим, прежде всего, простейшие свойства делимости многочленов.

1°. *Всякий многочлен $f(x)\neq 0$ из $P[x]$ делится на самого себя.*

Действительно, мы можем написать очевидное равенство

$$f(x)=f(x)\cdot 1,$$

а единицу 1 поля P можно рассматривать как многочлен нулевой степени из $P[x]$.

2°. *Если $f(x)$ и $g(x)$ — многочлены из $P[x]$, и $f(x)$ делится на $g(x)$, а $g(x)$ делится на $f(x)$, то многочлены $f(x)$ и $g(x)$ отличаются друг от друга лишь множителем нулевой степени:*

$$f(x)=cg(x) \qquad (c\neq 0),$$

где c — некоторый элемент из поля P.

В самом деле, так как $f(x)$ делится на $g(x)$, а $g(x)$ делится на $f(x)$, то по определению делимости мы можем написать, что

$$f(x)=g(x)q_1(x), \quad g(x)=f(x)q_2(x).$$

Подставляя выражение $g(x)$ из второго равенства в первое, получаем:

$$f(x)=f(x)q_1(x)q_2(x),$$

или, сокращая на $f(x)$ [1]):

$$1=q_1(x)q_2(x).$$

---

[1]) Законность такого сокращения основана на том, что $P[x]$ есть область целостности. В самом деле, если в некоторой области целостности имеет место равенство $ac=bc$ (или $ca=cb$), причём $c\neq 0$, то $ac-bc=0$ или $(a-b)c=0$. Отсюда, учитывая, что $c\neq 0$ и что кольцо не содержит делителей нуля, получаем: $a-b=0$, $a=b$.

В левой части последнего равенства находится 1, т. е. многочлен нулевой степени. Следовательно, для обеспечения равенства необходимо, чтобы произведение $q_1(x)q_2(x)$ было также многочленом нулевой степени, а это возможно лишь тогда, когда степени самих сомножителей $q_1(x)$ и $q_2(x)$ равны нулю. Таким образом, $q_1(x)=c$, $q_2(x)=d$, где c и d — элементы поля P, отличные от нуля. Отсюда $f(x)=cg(x)$, что и требовалось доказать.

В дальнейшем мы будем два многочлена, отличающихся друг от друга множителем нулевой степени, называть многочленами, совпадающими с точностью до множителя нулевой степени.

3°. *Если два многочлена* $f_1(x)$ *и* $f_2(x)$ *из* $P[x]$ *делятся на третий многочлен* $g(x)$ *из* $P[x]$, *то их сумма* $f_1(x)+f_2(x)$ *и разность* $f_1(x)-f_2(x)$ *делятся на* $g(x)$.

Доказательство проводится сходным образом. А именно, по определению делимости можно написать, что

$$f_1(x)=g(x)q_1(x),\quad f_2(x)=g(x)q_2(x),$$

где $q_1(x)$ и $q_2(x)$ — некоторые многочлены из $P[x]$. Складывая и вычитая почленно оба эти равенства, получаем:

$$f_1(x)\pm f_2(x)=g(x)q(x),$$

где $q(x)=q_1(x)\pm q_2(x)$ — многочлен из того же кольца $P[x]$. Мы видим отсюда, что $f_1(x)\pm f_2(x)$ делится на $g(x)$.

Свойство 3° (делимость) можно обобщить следующим образом:

4°. *Если многочлены* $f_1(x)$, $f_2(x)$, ..., $f_k(x)$ *из* $P[x]$ *делятся на многочлен* $g(x)$ *из* $P[x]$, *то* $c_1f_1(x)+c_2f_2(x)+\ldots+c_kf_k(x)$, *где* c_i — *произвольные элементы поля* P, *делится на* $g(x)$.

Доказательство этого свойства сходно с доказательством свойства 3°.

5°. *Если* $f_1(x)$, $f_2(x)$, $f_3(x)$, ..., $f_k(x)$ — *многочлены из* $P[x]$ *и* $f_1(x)$ *делится на многочлен* $g(x)$ *из* $P[x]$, *то произведение* $f_1(x)f_2(x)f_3(x)\ldots f_k(x)$ *делится на* $g(x)$.

В самом деле, если $f_1(x)$ делится на $g(x)$, то

$$f_1(x)=g(x)q_1(x), \tag{1}$$

где $q_1(x)$ — некоторый многочлен из $P[x]$. Умножая обе части равенства (1) на $f_2(x)f_3(x)\ldots f_k(x)$, получаем:

$$f_1(x)f_2(x)\ldots f_k(x)=g(x)q(x),$$

где $q(x)=q_1(x)f_2(x)\ldots f_k(x)$ — также многочлен из $P[x]$. Следовательно, $f_1(x)f_2(x)\ldots f_k(x)$ делится на $g(x)$.

6°. *Если* $f(x)$, $g(x)$ *и* $h(x)$ — *многочлены из* $P[x]$ *и* $f(x)$ *делится на* $g(x)$, *а* $g(x)$ *делится на* $h(x)$, *то* $f(x)$ *делится на* $h(x)$.

Для доказательства опять обращаемся к определению делимости многочленов. Пишем:

$$f(x)=g(x)q_1(x),\quad g(x)=h(x)q_2(x),$$

где $q_1(x)$ и $q_2(x)$ — некоторые многочлены из $P[x]$. Подставляя выражения $g(x)$ из второго равенства в первое, получаем:

$$f(x) = h(x)\,q(x),$$

где $q(x) = q_1(x)\,q_2(x)$ — многочлен из $P[x]$, т. е. $f(x)$ делится на $h(x)$.

Наконец, следует указать ещё одно свойство.

7°. *Многочлены нулевой степени из $P[x]$ являются делителями любого многочлена $f(x)$ из $P[x]$.*

Действительно, если $c \neq 0$ — элемент поля P и

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n$$

— какой-нибудь многочлен из $P[x]$, то, очевидно,

$$f(x) = c\left(\frac{a_0}{c} x^n + \frac{a_1}{c} x^{n-1} + \ldots + \frac{a_n}{c}\right) = c\,q(x),$$

где

$$q(x) = \frac{a_0}{c} x^n + \frac{a_1}{c} x^{n-1} + \ldots + \frac{a_n}{c}$$

есть снова многочлен из $P[x]$.

В кольце целых чисел имеют место аналогичные свойства делимости. При этом числа 1 и -1 играют роль, сходную с ролью многочленов нулевой степени. А именно, если целое число a делится на целое число b, а b делится на a, то числа a и b отличаются друг от друга лишь множителем ± 1. Далее, всякое целое число a делится на ± 1.

Делимость одного многочлена на другой можно обнаружить с помощью процесса, хорошо известного читателю из элементарной алгебры. Мы имеем в виду так называемый *алгорифм деления с остатком*. Но этот процесс нуждается в обосновании, так как заранее не очевидно, что он должен иметь место для любого поля P. Такое обоснование будет дано при доказательстве следующей теоремы:

Теорема 5 (о делении с остатком). *Для любых двух многочленов $f(x)$ и $g(x) \neq 0$ из $P[x]$ существует такая единственная пара многочленов $q(x)$ и $r(x)$ из того же кольца $P[x]$, что*

$$f(x) = g(x)\,q(x) + r(x) \tag{2}$$

и при $r(x) \neq 0$ степень $r(x)$ меньше степени $g(x)$.

Замечание. Обычно многочлен $q(x)$ называется *частным*, а многочлен $r(x)$ — *остатком* от деления $f(x)$ на $g(x)$.

Доказательство. Пусть

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n \qquad (a_0 \neq 0),$$

$$g(x) = b_0 x^m + b_1 x^{m-1} + \ldots + b_m \qquad (b_0 \neq 0).$$

Если $n < m$, то равенство (2) будет удовлетворяться при $q(x) = 0$, $r(x) = f(x)$.

Если $n \geqslant m$, то действуем следующим образом. Вычитаем из $f(x)$ многочлен $g(x)$, умноженный на $\frac{a_0}{b_0} x^{n-m}$:

$$f(x) - \frac{a_0}{b_0} x^{n-m} g(x) = f_1(x).$$

В результате старший член $a_0 x^n$ многочлена $f(x)$ уничтожится и степень $f(x)$ понизится:

$$f_1(x) = a_0' x^{n_1} + a_1' x^{n_1-1} + \ldots + a'_{n_1} \qquad (a_0' \neq 0,\ n_1 < n).$$

Если степень $f_1(x)$ больше или равна степени $g(x)$, то мы снова повторяем процесс понижения степени:

$$f_1(x) - \frac{a_0'}{b_0} x^{n_1-m} g(x) = f_2(x)$$

и т. д. Так как степени n, n_1, $n_2, \ldots$ не могут убывать безгранично, то в конце концов мы придём к многочлену $r(x)$, у которого степень будет ниже степени $g(x)$. Таким образом,

$$f(x) - \frac{a_0}{b_0} x^{n-m} g(x) = f_1(x),$$

$$f_1(x) - \frac{a_0'}{b_0} x^{n_1-m} g(x) = f_2(x),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$f_k(x) - \frac{a_0^{(k)}}{b_0} x^{n_k-m} g(x) = r(x).$$

Складывая эти равенства почленно, мы после очевидных упрощений получим:

$$f(x) - \left(\frac{a_0}{b_0} x^{n-m} + \frac{a_0'}{b_0} x^{n_1-m} + \ldots + \frac{a_0^{(k)}}{b_0} x^{n_k-m} \right) g(x) = r(x)$$

или

$$f(x) = g(x) q(x) + r(x),$$

где

$$q(x) = \frac{a_0}{b_0} x^{n-m} + \frac{a_0'}{b_0} x^{n_1-m} + \ldots + \frac{a_0^{(k)}}{b_0} x^{n_k-m}.$$

Коэффициенты многочленов $q(x)$ и $r(x)$ будут при этом принадлежать полю P, так как мы их получили с помощью операций сложения, вычитания, умножения и деления, не выводящих за пределы поля.

Для завершения доказательства остаётся убедиться, что частное и остаток определяются единственным образом.

Пусть, кроме $q(x)$ и $r(x)$, существует частное $q_1(x)$ и остаток $r_1(x)$. Тогда

$$f(x)=g(x)\,q_1(x)+r_1(x). \tag{3}$$

Из равенств (2) и (3) следует, что

$$g(x)\,q(x)+r(x)=g(x)\,q_1(x)+r_1(x)$$

или

$$g(x)\,[q(x)-q_1(x)]=r_1(x)-r(x). \tag{4}$$

Если $q(x)\neq q_1(x)$, то $q(x)-q_1(x)\neq 0$, вследствие чего и $r_1(x)-r(x)\neq 0$. Но тогда мы приходим к абсурду — в правой части равенства (4) находится разность $r_1(x)-r(x)$ со степенью, меньшей, чем m, так как степени $r_1(x)$ и $r(x)$ меньше m; в левой же части того же равенства мы имеем произведение $g(x)\,[q(x)-q_1(x)]$ со степенью не ниже m. Следовательно, $q(x)=q_1(x)$ и $r(x)=r_1(x)$.

Процесс, с помощью которого мы получили в только что изложенном доказательстве частное и остаток, является не чем иным, как правилом деления расположенных многочленов, известным ещё из школьного курса алгебры. Таким образом, доказывая теорему 5, мы попутно дали и обоснование этого правила.

С некоторым ограничением теорему 5 можно распространить и на случай произвольного коммутативного кольца R с единицей $e\neq 0$. А именно, справедлива следующая

Теорема 6. *Если R — коммутативное кольцо с единицей $e\neq 0$, $f(x)$ и $g(x)\neq 0$ — два многочлена из $R[x]$, причём старший коэффициент $g(x)$ равен единице, то существует такая единственная пара многочленов $q(x)$ и $r(x)$ из того же кольца $R[x]$, что*

$$f(x)=g(x)\,q(x)+r(x)$$

и при $r(x)\neq 0$ степень $r(x)$ меньше степени $g(x)$.

Доказательство ничем существенным не отличается от доказательства теоремы 5, и самый процесс нахождения частного и остатка здесь проще в том отношении, что вместо $\frac{a_0}{b_0}x^{n-m}, \frac{a_0'}{b_0}x^{n_1-m}$ и т. д. берутся в качестве множителей при $g(x)$ выражения a_0x^{n-m}, $a_0'x^{n_1-m},\ldots$ При доказательстве единственности частного и остатка надо учесть, что единица e кольца R не может быть делителем нуля [1]), в силу чего степень произведения $g(x)\,[q(x)-q_1(x)]$ должна быть в точности равна сумме степеней $g(x)$ и $q(x)-q_1(x)$.

Алгорифм деления с остатком позволяет обнаружить, делится ли данный многочлен $f(x)$ из $P[x]$ на многочлен $g(x)$ из $P[x]$ или

---

[1]) Если бы единица e была делителем нуля, то существовало бы такое $a\neq 0$, что $ae\neq 0$. Но это противоречит равенству $ae=a\neq 0$.

не делится. Именно, $f(x)$ *делится на* $g(x)$ *тогда и только тогда, когда остаток* $r(x)$ *от деления* $f(x)$ *на* $g(x)$ *равен нулю.*

Действительно, если остаток $r(x)$ равен нулю, то равенство (2) превращается в

$$f(x)=g(x)\,q(x),$$

откуда ясно, что $f(x)$ делится на $g(x)$. Обратно, если $f(x)$ делится на $g(x)$, то

$$f(x)=g(x)\,q(x),$$

где $q(x)$ — некоторый многочлен из $P[x]$. Отсюда в силу единственности остатка и частного следует, что остаток $r(x)$ равен нулю.

Благодаря этой связи делимости с равенством нулю остатка получается, что делимость многочлена $f(x)$ на многочлен $g(x)$ не зависит от того, над каким полем рассматриваются многочлены $f(x)$ и $g(x)$. В самом деле, будем ли мы рассматривать поле P или более обширное поле, содержащее P как часть, мы получим при делении $f(x)$ на $g(x)$ одни и те же частное и остаток.

Роль алгорифма деления с остатком этим, однако, не исчерпывается. Мы сейчас увидим, что на основании теоремы о делении с остатком можно провести дальнейший параллелизм между теорией делимости целых чисел и теорией делимости многочленов. К тому же в кольце целых чисел, как известно, имеет место теорема, аналогичная теореме 5.

Пусть $f(x)$ и $g(x)$ — два каких-нибудь многочлена из $P[x]$. Назовём третий многочлен $d(x)$ из того же кольца $P(x)$ общим делителем $f(x)$ и $g(x)$, если $d(x)$ делит как $f(x)$, так и $g(x)$. В частности, общий делитель $D(x)$ называется наибольшим, если $D(x)$ делится на всякий общий делитель $d(x)$ многочленов $f(x)$ и $g(x)$.

Мы покажем, что для любых двух многочленов $f(x)$ и $g(x)$ из $P[x]$ наибольший общий делитель существует, а именно укажем вполне определённый способ, позволяющий находить наибольший общий делитель для каких угодно многочленов $f(x)$ и $g(x)\neq 0$ из $P[x]$. Этот способ, известный под названием *алгорифма Евклида*, заключается в следующем. Пусть степень $f(x)$ не ниже степени $g(x)$. Тогда делим $f(x)$ на $g(x)$; остаток и частное, полученные при делении, обозначим, соответственно, через $r_1(x)$ и $q_1(x)$. Затем делим $g(x)$ на остаток $r_1(x)$; в результате получатся второй остаток $r_2(x)$ и частное $q_2(x)$ и т. д. Вообще каждый раз делится предыдущий остаток на последующий. Степени получающихся при этом процессе остатков $r_1(x)$, $r_2(x)$, ... будут, очевидно, всё время убывать. Но целые неотрицательные числа не могут убывать неограниченно. Следовательно, этот процесс деления не может быть бесконечным — в конце концов мы должны притти к остатку $r_k(x)$, на который нацело разделится предыдущий остаток $r_{k-1}(x)$. Покажем,

что *этот последний остаток* $r_k(x)$ *и будет наибольшим общим делителем многочленов* $f(x)$ *и* $g(x)$.

Запишем весь процесс деления следующим образом:

$$\left.\begin{aligned} f(x) &= g(x)\,q_1(x) + r_1(x),\\ g(x) &= r_1(x)\,q_2(x) + r_2(x),\\ &\cdots\cdots\cdots\cdots\cdots\\ r_{k-2}(x) &= r_{k-1}(x)\,q_k(x) + r_k(x).\\ r_{k-1}(x) &= r_k(x)\,q_{k+1}(x). \end{aligned}\right\} \qquad (5)$$

Прежде всего покажем, что $r_k(x)$ есть общий делитель многочленов $f(x)$ и $g(x)$. Обратимся к предпоследнему равенству системы равенств (5):

$$r_{k-2}(x) = r_{k-1}(x)\,q_k(x) + r_k(x).$$

Его правая часть делится на $r_k(x)$, так как $r_{k-1}(x)$ делится на $r_k(x)$, а $r_k(x)$ делится на самого себя. Следовательно, левая часть также делится на $r_k(x)$, т. е. $r_{k-2}(x)$ делится на $r_k(x)$. Обращаемся, далее, к вышележащему равенству

$$r_{k-3}(x) = r_{k-2}(x)\,q_{k-1}(x) + r_{k-1}(x).$$

Здесь $r_{k-2}(x)$ и $r_{k-1}(x)$ делятся на $r_k(x)$, откуда ясно, что вся правая часть делится на $r_k(x)$. Следовательно, на $r_k(x)$ делится и левая часть, т. е. $r_{k-3}(x)$ делится на $r_k(x)$. Двигаясь таким образом постепенно вверх, мы, наконец, дойдём до многочленов $g(x)$ и $f(x)$ и убедимся, что $g(x)$ и $f(x)$ делятся на $r_k(x)$.

Теперь остаётся показать, что $r_k(x)$ есть наибольший общий делитель. Для этой цели обратимся к первому равенству

$$f(x) = g(x)\,q_1(x) + r_1(x)$$

и посмотрим, что получается относительно некоторого общего делителя $d(x)$. Так как $f(x)$ и $g(x)$ делятся на $d(x)$, то разность $f(x) - g(x)\,q_1(x) = r_1(x)$ должна делиться на $d(x)$. Точно так же, рассматривая второе из равенств системы (5)

$$g(x) = r_1(x)\,q_2(x) + r_2(x),$$

находим, что $r_2(x)$ делится на $d(x)$ и т. д. Так, опускаясь постепенно вниз, мы, наконец, дойдём до $r_k(x)$ и убедимся, что $r_k(x)$ делится на $d(x)$. Иными словами, мы обнаружим, что $r_k(x)$ есть наибольший общий делитель многочленов $f(x)$ и $g(x)$.

Нетрудно убедиться, что *наибольший общий делитель многочленов* $f(x)$ *и* $g(x)$ *является единственным с точностью до множителя нулевой степени.*

В самом деле, если $D_1(x)$ и $D_2(x)$ — два наибольших общих делителя многочленов $f(x)$ и $g(x)$, то по определению наибольшего общего делителя $D_1(x)$ должно делиться на $D_2(x)$ и $D_2(x)$ должно

делиться на $D_1(x)$, откуда по свойству 2° делимости $D_2(x) = cD_1(x)$, что и требовалось показать.

Наибольший общий делитель $f(x)$ и $g(x)$ может оказаться многочленом нулевой степени. В этом случае $f(x)$ и $g(x)$ называются *взаимно простыми многочленами.*

Как и в случае целых чисел, условимся наибольший общий делитель многочленов $f(x)$ и $g(x)$ обозначать для сокращения письма символом $(f(x), g(x))$.

Пример 1. Найти наибольший общий делитель многочленов

$$f(x) = 2x^5 - 3x^4 - 5x^3 + x^2 + 6x + 3,$$
$$g(x) = 3x^4 + 2x^3 - 3x^2 - 5x - 2$$

над полем рациональных чисел. Чтобы избежать дробных коэффициентов, умножим предварительно $f(x)$ на 3:

$$\begin{array}{r|l} 6x^5 - 9x^4 - 15x^3 + 3x^2 + 18x + 9 & 3x^4 + 2x^3 - 3x^2 - 5x - 2 \\ \underline{6x^5 + 4x^4 - 6x^3 - 10x^2 - 4x} & \overline{2x} \\ -13x^4 - 9x^3 + 13x^2 + 22x + 9 & \end{array}$$

Теперь, чтобы избежать дробных коэффициентов, умножим полученную разность на 3. Этим мы, правда, исказим частное, но остаток определится с точностью до множителя нулевой степени. Итак, продолжаем вычисления:

$$\begin{array}{r|l} -39x^4 - 27x^3 + 39x^2 + 66x + 27 & 3x^4 + 2x^3 - 3x^2 - 5x - 2 \\ \underline{-39x^4 - 26x^3 + 39x^2 + 65x + 26} & \overline{-13} \\ -x^3 + x + 1 & \end{array}$$

Таким образом, мы нашли с точностью до множителя нулевой степени остаток

$$r_1(x) = x^3 - x - 1$$

от деления $f(x)$ на $g(x)$. Теперь надо $g(x)$ делить на $r_1(x)$. Читатель может сам без труда убедиться, что $g(x)$ делится без остатка на $r_1(x)$. Следовательно,

$$x^3 - x - 1$$

и есть наибольший общий делитель многочленов $f(x)$ и $g(x)$.

Так как алгорифм Евклида сводится к последовательному применению алгорифма деления с остатком, то можно высказать следующее важное заключение: наибольший общий делитель $D(x)$ многочленов $f(x)$ и $g(x)$, найденный с помощью алгорифма Евклида, не зависит от того, будем ли мы рассматривать $f(x)$ и $g(x)$ над полем P или над более обширным полем P'.

Так, в только что разобранном примере мы нашли, что наибольший общий делитель многочленов

$$f(x) = 2x^5 - 3x^4 - 5x^3 + x^2 + 6x + 3,$$
$$g(x) = 3x^4 + 2x^3 - 3x^2 - 5x - 2$$

над полем рациональных чисел равен $x^3 - x - 1$. Но эти же многочлены и над полем действительных чисел будут иметь $x^3 - x - 1$ наибольшим общим делителем.

Для нескольких многочленов $f_1(x)$, $f_2(x)$, ..., $f_k(x)$ из $P[x]$ общий делитель и наибольший общий делитель определяются аналогичным образом. Именно, многочлен $d(x)$ из $P[x]$ называется *общим делителем* $f_1(x)$, $f_2(x)$, ..., $f_k(x)$, если каждый из многочленов $f_1(x)$, $f_2(x)$, ..., $f_k(x)$ делится на $d(x)$. Общий делитель $D(x)$ называется наибольшим, если $D(x)$ делится на всякий общий делитель многочленов $f_1(x)$, $f_2(x)$, ..., $f_k(x)$.

Разыскание наибольшего общего делителя нескольких многочленов можно свести к нахождению наибольшего общего делителя двух многочленов. В самом деле, пусть $D_1(x)$ — наибольший общий делитель $k-1$ многочленов $f_1(x)$, $f_2(x)$, ..., $f_{k-1}(x)$. Легко убедиться, что если $D(x)$ — наибольший общий делитель $D_1(x)$ и $f_k(x)$, то это $D(x)$ будет вместе с тем и наибольшим общим делителем всех k многочленов $f_1(x)$, ..., $f_k(x)$.

Повторяя дословно те же рассуждения, что и выше, можно убедиться в единственности наибольшего общего делителя нескольких многочленов (с точностью до множителя нулевой степени).

Исходя из алгорифма Евклида, можно получить ряд выводов. Отметим наиболее существенное.

Теорема 7. *Если $D(x)$ — наибольший общий делитель многочленов $f(x)$ и $g(x)$ из $P[x]$, то в том же самом кольце $P[x]$ можно подобрать такую пару многочленов $\varphi(x)$ и $\psi(x)$, что*

$$f(x)\varphi(x) + g(x)\psi(x) = D(x). \tag{6}$$

Доказательство. Возьмём предпоследнее равенство (5) и перенесём $r_{k-1}(x)q_k(x)$ в левую часть. Тогда, принимая во внимание, что $r_k(x) = D(x)$, получаем:

$$r_{k-2}(x) - r_{k-1}(x)q_k(x) = D(x). \tag{7}$$

Затем из равенства

$$r_{k-3}(x) = r_{k-2}(x)q_{k-1}(x) + r_{k-1}(x)$$

определяем $r_{k-1}(x)$

$$r_{k-1}(x) = r_{k-3}(x) - r_{k-2}(x)q_{k-1}(x)$$

и подставляем это значение $r_{k-1}(x)$ в (7). Получим:

$$r_{k-2}(x)[1 + q_k(x)q_{k-1}(x)] - r_{k-3}(x)q_k(x) = D(x)$$

или

$$r_{k-2}(x)\varphi_1(x)+r_{k-3}(x)\psi_1(x)=D(x), \tag{8}$$

где

$$\varphi_1(x)=1+q_k(x)q_{k-1}(x), \quad \psi_1(x)=-q_k(x).$$

Далее, из равенства

$$r_{k-4}(x)=r_{k-3}(x)q_{k-2}(x)+r_{k-2}(x)$$

определяем $r_{k-2}(x)$ и подставляем в (8). Получим:

$$r_{k-3}(x)\varphi_2(x)+r_{k-4}(x)\psi_2(x)=D(x)$$

и т. д., пока не получится равенство

$$f(x)\varphi_{k-2}(x)+g(x)\psi_{k-2}(x)=D(x),$$

т. е. равенство (6) с $\varphi(x)=\varphi_{k-2}(x)$ и $\psi(x)=\psi_{k-2}(x)$.

В частности, когда многочлены $f(x)$ и $g(x)$ взаимно просты, равенство (6) принимает вид

$$f(x)\varphi(x)+g(x)\psi(x)=c \qquad (c\neq 0).$$

Мы можем c положить равным 1, так как обе части последнего равенства можно разделить на c и в качестве $\varphi(x)$ и $\psi(x)$ рассматривать, соответственно,

$$\frac{\varphi(x)}{c} \quad \text{и} \quad \frac{\psi(x)}{c}.$$

Таким образом, если многочлены $f(x)$ и $g(x)$ взаимно просты, то в том же кольце $P[x]$ можно подобрать такую пару многочленов $\varphi(x)$ и $\psi(x)$, что

$$f(x)\varphi(x)+g(x)\psi(x)=1. \tag{6*}$$

Пример 2. Для многочленов

$$f(x)=2x^5-3x^4-5x^3+x^2+6x+3,$$
$$g(x)=3x^4+2x^3-3x^2-5x-2$$

над полем рациональных чисел подобрать многочлены $\varphi(x)$ и $\psi(x)$ над тем же полем так, чтобы

$$f(x)\varphi(x)+g(x)\psi(x)=D(x).$$

Здесь существенное значение имеют не только остатки, но и частные, получаемые в процессе последовательного деления; поэтому необходимо каждый раз учитывать производимое сокращение или умножение на число. С данными многочленами $f(x)$ и $g(x)$ мы уже имели дело в примере 1. Учитывая умножение $f(x)$ на 3, а также умножение на 3 многочлена $-13x^4-9x^3+13x^2+22x+9$, мы

можем соответствующие результаты вычислений, полученные в предыдущем примере, записать коротко в виде равенств:

$$3f(x)=g(x)\cdot 2x+(-13x^4-9x^3+13x^2+22x+9), \quad (9)$$

$$3(-13x^4-9x^3+13x^2+22x+9)=g(x)\cdot(-13)-r_1(x), \quad (10)$$

где

$$r_1(x)=x^3-x-1.$$

Мы знаем из предыдущего примера, что $r_1(x)$ есть наибольший общий делитель многочленов $f(x)$ и $g(x)$, т. е. $r_1(x)=D(x)$.

Умножим обе части равенства (9) на 3 и затем подставим значение

$$3(-13x^4-9x^3+13x^2+22x+9)$$

из равенства (10). Получаем:

$$9f(x)=g(x)\cdot 6x+g(x)\cdot(-13)-r_1(x),$$

или

$$9f(x)=g(x)(6x-13)-D(x).$$

Отсюда

$$f(x)\cdot(-9)+g(x)\cdot(6x-13)=D(x),$$

т. е. мы нашли, что

$$\varphi(x)=-9, \quad \psi(x)=6x-13.$$

Пользуясь равенством (6*), можно без труда получить ряд свойств взаимно простых многочленов, аналогичных свойствам взаимно простых целых чисел.

1°. *Два многочлена $f(x)$ и $g(x)$ из $P[x]$ тогда и только тогда взаимно просты, когда*

$$f(x)\varphi(x)+g(x)\psi(x)=1, \quad (6^*)$$

где $\varphi(x)$ и $\psi(x)$ — некоторые многочлены из $P[x]$.

Доказательство. Если $f(x)$ и $g(x)$ взаимно просты, то, как мы уже знаем, должно выполняться равенство (6*) для некоторых $\varphi(x)$ и $\psi(x)$ из $P[x]$.

Обратно, пусть имеет место равенство (6*) для некоторых $\varphi(x)$ и $\psi(x)$ из $P[x]$. Обозначим через $d(x)$ произвольный общий делитель $f(x)$ и $g(x)$. Тогда, очевидно, левая часть равенства (6*) будет делиться на $d(x)$, в силу чего на $d(x)$ будет делиться и правая часть, т. е. 1. Но $d(x)$ может делить 1 только в том случае, когда $d(x)$ есть многочлен нулевой степени. Итак $f(x)$ и $g(x)$ могут иметь общим делителем только многочлен нулевой степени, откуда $f(x)$ и $g(x)$ взаимно просты.

2°. *Если $D(x)$ — наибольший общий делитель многочленов $f(x)$ и $g(x)$ из $P[x]$, то многочлены $f_1(x)$ и $g_1(x)$, получающиеся при делении $f(x)$ и $g(x)$ на $D(x)$, будут взаимно простыми.*

Доказательство. Так как по условию

$$f(x)=f_1(x)D(x),\quad g(x)=g_1(x)D(x),$$

то равенство

$$f(x)\varphi(x)+g(x)\psi(x)=D(x)$$

можно переписать следующим образом:

$$f_1(x)\varphi(x)D(x)+g_1(x)\psi(x)D(x)=D(x).$$

Сокращая на $D(x)$, получаем:

$$f_1(x)\varphi(x)+g_1(x)\psi(x)=1.$$

Отсюда в силу предыдущего свойства 1° следует, что $f_1(x)$ и $g_1(x)$ взаимно просты.

3°. *Если многочлены* $f(x)$ *и* $g(x)$ *из* $P[x]$ *взаимно просты с третьим многочленом* $h(x)$ *из* $P[x]$, *то произведение* $f(x)g(x)$ *также взаимно просто с* $h(x)$.

Доказательство. Так как по условию $f(x)$ и $h(x)$ взаимно просты, то

$$f(x)\varphi(x)+h(x)\psi(x)=1$$

для некоторых $\varphi(x)$ и $\psi(x)$ из $P[x]$. Умножим обе части этого равенства на $g(x)$:

$$f(x)g(x)\varphi(x)+h(x)g(x)\psi(x)=g(x).$$

Пусть теперь $d(x)$ — какой-нибудь общий делитель $f(x)g(x)$ и $h(x)$. Тогда левая часть последнего равенства будет делиться на $d(x)$ и потому на $d(x)$ будет делиться и правая часть, т. е. $g(x)$. Таким образом, $d(x)$ оказалось общим делителем $g(x)$ и $h(x)$. Но $g(x)$ и $h(x)$ взаимно просты. Следовательно, $d(x)$ есть многочлен нулевой степени. Тем самым мы показали, что произведение $f(x)g(x)$ взаимно просто с $h(x)$.

Свойство 3° можно с помощью метода математической индукции обобщить на любое число многочленов из $P[x]$: *если каждый из многочленов* $f_1(x)$, $f_2(x)$, ..., $f_k(x)$ *взаимно прост с* $h(x)$, *то произведение* $f_1(x)f_2(x)\ldots f_k(x)$ *также взаимно просто с* $h(x)$.

4°. *Если* $f(x)$, $g(x)$, $h(x)$ — *такие многочлены из* $P[x]$, *что* $f(x)$ *взаимно просто с* $h(x)$ *и произведение* $f(x)g(x)$ *делится на* $h(x)$, *то* $g(x)$ *делится на* $h(x)$.

Доказательство. Так как $f(x)$ и $h(x)$ взаимно просты, то

$$f(x)\varphi(x)+h(x)\psi(x)=1$$

для некоторых $\varphi(x)$ и $\psi(x)$ из $P[x]$. Умножим обе части этого равенства на $g(x)$. Получим:

$$f(x)g(x)\varphi(x)+h(x)g(x)\psi(x)=g(x).$$

Левая часть последнего равенства делится на $h(x)$; следовательно, правая часть, т. е. $g(x)$ делится на $h(x)$.

В теории делимости многочленов роль простого числа играют так называемые неприводимые многочлены.

Определение. *Многочлен $f(x)$ из $P[x]$ называется приводимым в поле P, если он может быть разложен в произведение двух многочленов меньшей степени из того же кольца $P[x]$.*

Напротив, многочлен $p(x)$ из $P(x)$ выше нулевой степени называется неприводимым в поле P, если $p(x)$ не может быть разложен в произведение двух многочленов меньшей степени из того же кольца $P[x]$.

Согласно этому определению многочлен нулевой степени нельзя считать приводимым, а также нельзя считать неприводимым многочленом. В этом отношении наблюдается та же картина, что и для числа 1. Число 1, как известно, не считается простым и в то же время не считается и составным числом.

Пример 3. Рассмотрим многочлен

$$f(x) = x^4 - 5x^2 + 6$$

над полем рациональных чисел. Он разлагается в произведение многочленов меньшей (а именно второй) степени над тем же полем рациональных чисел:

$$f(x) = (x^2 - 2)(x^2 - 3).$$

Следовательно, рассматриваемый многочлен $f(x)$ приводим в поле рациональных чисел.

Пример 4. Многочлен первой степени

$$p(x) = x + 1$$

над произвольным полем P неприводим в P.

Действительно, если $f(x)$ и $g(x)$ — произвольные многочлены выше нулевой степени, то их произведение будет иметь по меньшей мере вторую, а не первую степень.

Пример 5. Многочлен

$$p(x) = x^3 - 2$$

неприводим в поле рациональных чисел.

В самом деле, если бы многочен $p(x)$ был приводим в поле рациональных чисел, то $p(x)$ разлагался бы в произведение двух множителей, из которых один был бы первой степени, а другой — второй степени:

$$p(x) = x^3 - 2 = (ax + b)(cx^2 + dx + e),$$

где a, b, c, d, e — некоторые рациональные числа. Полагая $x=-\frac{b}{a}$, получаем:

$$-\frac{b^3}{a^3}-2=0$$

или

$$\sqrt[3]{2}=-\frac{b}{a}.$$

Получился абсурд: $\sqrt[3]{2}$ оказался рациональным числом $-\frac{b}{a}$. Следовательно, многочлен $p(x)=x^3-2$ неприводим в поле рациональных чисел.

Однако этот же многочлен в поле действительных чисел будет уже приводимым — так как в поле действительных чисел мы считаем допустимым разложение и на множители с иррациональными коэффициентами, то мы можем написать, что

$$p(x)=(x-\sqrt[3]{2})(x^2+\sqrt[3]{2}x+\sqrt[3]{4}).$$

Для многочленов из кольца $P[x]$ имеет место теорема, аналогичная теореме о разложении целого числа на простые множители.

Теорема 8. *Всякий многочлен из $P[x]$ выше нулевой степени разлагается в произведение неприводимых многочленов:*

$$f(x)=p_1(x)p_2(x)\ \ldots\ p_r(x)$$

($p_i(x)$ — неприводимый многочлен в поле P), и это разложение является единственным с точностью до порядка следования и множителей нулевой степени.

Для доказательства этой теоремы придётся предварительно рассмотреть следующие свойства неприводимых многочленов, сходные со свойствами простого числа.

1°. *Если $p_1(x)$ и $p_2(x)$ — неприводимые многочлены в поле P и $p_1(x)$ делится на $p_2(x)$, то $p_1(x)$ и $p_2(x)$ совпадают с точностью до множителя нулевой степени.*

В самом деле, из равенства $p_1(x)=p_2(x)q(x)$, где $q(x)$ — частное от деления $p_1(x)$ на $p_2(x)$, следует в силу неприводимости $p_1(x)$, что $q(x)$ есть многочлен нулевой степени: $q(x)=c\neq 0$. Отсюда $p_1(x)=cp_2(x)$, что и требовалось показать.

2°. *Многочлен $f(x)$ из $P[x]$ тогда и только тогда не делится на многочлен $p(x)$, неприводимый в поле P, когда $f(x)$ и $p(x)$ взаимно просты.*

Доказательство. Пусть $f(x)$ не делится на $p(x)$. Обозначим через $D(x)$ наибольший общий делитель $f(x)$ и $p(x)$. Так как $p(x)$ — неприводимый многочлен, то из условия делимости $p(x)$ на $D(x)$ следует лишь одно из двух: либо 1) $D(x)$ есть многочлен нулевой степени, либо 2) $D(x)$ совпадает с $p(x)$ с точностью до

множителя нулевой степени. Вторая возможность, однако, отпадает, так как в случае совпадения $D(x)$ с $p(x)$ (с точностью до множителя нулевой степени) многочлен $f(x)$ делился бы на $p(x)$. Следовательно, остаётся только одно — $D(x)$ является многочленом нулевой степени. Но это значит, что $f(x)$ и $p(x)$ взаимно просты.

Обратно, пусть $f(x)$ и $p(x)$ взаимно просты. Тогда $f(x)$ не может делиться на $p(x)$: если бы $f(x)$ делилось на $p(x)$, то наибольший общий делитель $f(x)$ и $p(x)$ был бы равен $p(x)$, а не многочлену нулевой степени.

3°. *Если произведение $f(x)g(x)$ двух многочленов из $P[x]$ делится на многочлен $p(x)$, неприводимый в P, то на $p(x)$ делится по меньшей мере один из сомножителей $f(x)$, $g(x)$.*

Доказательство. Предположим противное — пусть ни $f(x)$ и ни $g(x)$ не делятся на $p(x)$. Тогда по предыдущему свойству 2° многочлены $f(x)$ и $g(x)$ будут взаимно просты с $p(x)$. Отсюда в силу свойства 3° взаимно простых многочленов произведение $f(x)g(x)$ будет также взаимно просто с $p(x)$ и потому не может делиться на $p(x)$, что противоречит условию.

Очевидно, что свойство 3° можно распространить на случай произведения любого числа сомножителей, стоит только воспользоваться методом математической индукции.

Теперь мы можем приступить к доказательству теоремы.

Доказательство теоремы 8. Покажем сначала, что всякий многочлен $f(x)$ из $P[x]$ выше нулевой степени можно разложить в произведение неприводимых множителей.

Для неприводимого $f(x)$ утверждение очевидно — в этом случае получается разложение из одного неприводимого множителя: $f(x)=f(x)$. Поэтому пусть $f(x)$ приводимо. Тогда

$$f(x)=f_1(x)f_2(x),$$

где $f_1(x)$ и $f_2(x)$ — многочлены из $P[x]$ более низкой степени, чем $f(x)$. Если один или оба сомножителя $f_1(x)$ и $f_2(x)$ приводимы, то один или оба сомножителя $f_1(x)$ и $f_2(x)$ будут разлагаться на дальнейшие сомножители ещё более низкой степени и т. д. Этот процесс дальнейшего разложения на множители не может быть безграничным, так как степени многочленов не могут безгранично понижаться. Следовательно, мы в конце концов дойдём до разложения многочлена $f(x)$ на неприводимые множители.

Теперь остаётся доказать вторую половину теоремы — единственность разложения на неприводимые множители.

Пусть многочлен $f(x)$ двумя способами разлагается в произведение неприводимых множителей:

$$f(x)=p_1(x)p_2(x)\ldots p_k(x), \tag{11}$$

где $p_i(x)$ — многочлены, неприводимые в поле P, и

$$f(x)=q_1(x)q_2(x)\ldots q_l(x), \tag{12}$$

где $q_i(x)$ — также неприводимые в P многочлены. Без ограничения общности можно предположить, что $k \leqslant l$.

Из равенств (12) и (13) следует, что

$$p_1(x)p_2(x) \ldots p_k(x) = q_1(x)q_2(x) \ldots q_l(x). \qquad (13)$$

Левая часть последнего равенства делится, очевидно, на $p_1(x)$; следовательно, на $p_1(x)$ должна делиться и правая часть. Отсюда в силу свойства 3° неприводимого многочлена должен делиться на $p_1(x)$ по меньшей мере один из сомножителей правой части. Пусть для определённости $q_1(x)$ делится на $p_1(x)$. Тогда по свойству 1° неприводимого многочлена $q_1(x)$ и $p_1(x)$ должны совпадать с точностью до множителя нулевой степени: $q_1(x) = c_1 p_1(x)$. Подставляя это значение $q_1(x)$ в правую часть равенства (13) и производя сокращение обеих частей равенства на $p_1(x)$, получаем:

$$p_2(x) \ldots p_k(x) = c_1 q_2(x) \ldots q_l(x). \qquad (14)$$

Повторяем относительно равенства (14) аналогичные рассуждения. Получим $q_2(x) = c_2 p_2(x)$ и затем после соответствующего сокращения

$$p_3(x) \ldots p_k(x) = c_1 c_2 q_3(x) \ldots q_l(x)$$

и т. д. Мы утверждаем теперь, что $k = l$. В самом деле, если бы $k < l$, то после всех таких последовательных сокращений мы получили бы равенство

$$1 = c_1 c_2 \ldots c_k q_{k+1}(x) \ldots q_l(x).$$

Но это равенство абсурдно, так как 1 не может делиться на многочлены $q_{k+1}(x), \ldots, q_l(x)$, имеющие степень выше нулевой.

Итак, $k = l$ и $q_1(x) = c_1 p_1(x), \ldots, q_l(x) = c_l p_l(x)$. Теорема полностью доказана.

В разложении многочлена $f(x)$ на неприводимые множители могут встречаться многочлены, совпадающие с точностью до множителя нулевой степени. Например, многочлен

$$f(x) = 6(x^3 - 1)^2$$

в поле рациональных чисел разлагается в произведение четырёх неприводимых многочленов:

$$f(x) = (2x^2 + 2x + 2)(x^2 + x + 1)(x - 1)(3x - 3),$$

и мы видим, что многочлены

$$2x^2 + 2x + 2 \quad \text{и} \quad x^2 + x + 1$$

совпадают с точностью до множителя 2, а многочлены

$$x - 1 \quad \text{и} \quad 3x - 3$$

совпадают с точностью до множителя 3.

Пусть в разложении

$$f(x)=p_1(x)p_2(x)\dots p_k(x)$$

многочлена $f(x)$ на неприводимые многочлены в поле P многочлены $p_1(x)$, $p_2(x)$ и т. д. повторяются (с точностью до множителя нулевой степени) соответственно α_1 раз, α_2 раз и т. д. Тогда, объединяя повторяющиеся многочлены, получим так называемое каноническое разложение $f(x)$ в произведение неприводимых многочленов:

$$f(x)=cp_1^{\alpha_1}(x)p_2^{\alpha_2}(x)\dots p_r^{\alpha_r}(x) \qquad (r\leqslant k,\ c\neq 0 \text{ — элемент из } P).$$

Здесь неприводимые многочлены $p_i(x)$ между собой существенно различны (т. е. не совпадают с точностью до множителя нулевой степени).

Показатель α_i называется *кратностью* неприводимого множителя $p_i(x)$.

Так, например, рассмотренный выше многочлен

$$f(x)=6(x^3-1)^2$$

имеет следующее каноническое разложение:

$$f(x)=6(x^2+x+1)^2(x-1)^2,$$

причём кратности неприводимых множителей равны двум.

Вообще мы скажем, что некоторый многочлен $g(x)$ из $P[x]$ входит в данный многочлен $f(x)$ из $P[x]$ с кратностью α, если $f(x)$ делится на $g^{\alpha}(x)$, но не делится на $g^{\alpha+1}(x)$.

Пример 6. С какой кратностью входит многочлен $g(x)=x^2-4$ в многочлен $f(x)=x^5+x^4-8x^3-8x^2+16x+16$?

Применяя алгорифм деления с остатком, без труда убеждаемся, что $f(x)$ делится на $g^2(x)$, но не делится на $g^3(x)$. Следовательно, $g(x)$ входит в $f(x)$ с кратностью 2.

Пример 7. С какой кратностью входит многочлен $g(x)=x^2-2x-2$ в многочлен $f(x)=x^5-3x-3$?

Легко убедиться, что $f(x)$ не делится на $g(x)$. Это означает, что $g(x)$ входит в $f(x)$ с нулевой кратностью.

§ 3. Деление на линейный двучлен $x-a$. Корни многочленов

Ради большей общности выводов мы в этом параграфе будем рассматривать многочлены над произвольным коммутативным кольцом R с единицей $e\neq 0$. Очевидно, что поле является частным случаем такого кольца.

Мы займёмся часто встречающейся в алгебре задачей — задачей деления многочлена $f(x)$ из $R[x]$ на линейный двучлен $x-a$ с a, лежащим в том же кольце R, что и коэффициенты многочлена $f(x)$.

Так как старший коэффициент двучлена $x-a$ равен единице, то согласно теореме 6 мы можем написать, что

$$f(x)=(x-a)\,q(x)+r. \tag{1}$$

Очевидно, что остаток r должен быть некоторым элементом кольца R, так как если $r\neq 0$, то степень r должна быть ниже степени делителя $x-a$.

Равенство (1) остаётся в силе и при любом значении x [см. в § 1 соотношения (7)]. Возьмём для x значение a. Тогда

$$f(a)=(a-a)\,q(a)+r,$$

или, так как $a-a=0$,

$$f(a)=r.$$

Мы пришли к следующей теореме:

Теорема 9. *Остаток, получающийся при делении многочлена $f(x)$ над кольцом R на линейный двучлен $x-a$ над тем же кольцом R, равен значению многочлена при $x=a$.*

Пользуясь этой теоремой, можно находить остаток, не производя деления многочлена $f(x)$ на $x-a$.

Пример 1. Найти остаток от деления многочлена

$$f(x)=3x^4-x^3-2x^2-x+1$$

над кольцом целых чисел на $x+2$, не производя деления.

Так как $x+2=x-(-2)$, то здесь $a=-2$. Таким образом, по теореме 9 получаем следующий остаток:

$$r=f(-2)=3\cdot(-2)^4-(-2)^3-2\cdot(-2)^2-(-2)+1=$$
$$=48+8-8+2+1=51.$$

Деление многочлена $f(x)$ на линейный двучлен $x-a$ осуществляется особенно просто с помощью схемы Горнера, заключающейся в следующем.

Так как степень частного $q(x)$ от деления многочлена $f(x)$ на $x-a$ должна быть на единицу ниже, то мы можем положить:

$$q(x)=b_0x^{n-1}+b_1x^{n-2}+\ldots+b_{n-1}.$$

Подставляя выражения $f(x)$ и $q(x)$ в равенство (1), получаем:

$$a_0x^n+a_1x^{n-1}+\ldots+a_n=$$
$$=(x-a)\,(b_0x^{n-1}+b_1x^{n-2}+\ldots+b_{n-1})+r$$

или, производя в первой части перемножение и группируя по степеням x:

$$a_0x^n+a_1x^{n-1}+\ldots+a_n=$$
$$=b_0x^n+(b_1-ab_0)\,x^{n-1}+\ldots+(r-ab_{n-1}).$$

Отсюда согласно определению равенства двух многочленов следует, что

$$b_0 = a_0, \quad b_1 - ab_0 = a_1, \quad b_2 - ab_1 = a_2, \ \ldots, \quad b_{n-1} - ab_{n-2} = a_{n-1},$$
$$r - ab_{n-1} = a_n,$$

откуда

$$\left.\begin{aligned} b_0 &= a_0, \quad b_1 = ab_0 + a_1, \quad b_2 = ab_1 + a_2, \ \ldots \\ \ldots, \ b_{n-1} &= ab_{n-2} + a_{n-1}, \quad r = ab_{n-1} + a_n. \end{aligned}\right\} \tag{2}$$

Формулы (2) позволяют последовательно находить коэффициенты частного и остаток. Вычисления по формулам (2) удобнее всего проводить по следующей схеме, известной под названием схемы Горнера:

	a_0	a_1	a_2	$\ldots$	a_{n-1}	a_n
a	a_0	$ab_0 + a_1$	$ab_1 + a_2$	$\ldots$	$ab_{n-2} + a_{n-1}$	$ab_{n-1} + a_n$

В верхней строке схемы Горнера написаны в порядке убывания степеней x коэффициенты многочлена $f(x)$, а в нижней строке — коэффициенты b_i частного $q(x)$ и остаток r.

Поясним метод Горнера на нескольких примерах.

Пример 2. Пользуясь схемой Горнера, разделить многочлен

$$f(x) = 2x^5 - 5x^3 - 8x + 1$$

над кольцом целых чисел на $x - 3$.

Составляем схему Горнера. При этом надо выписывать все коэффициенты $f(x)$ без пропусков. Так, в данном многочлене отсутствуют члены с x^4 и x^2. Это значит, что $a_1 = 0$ и $a_3 = 0$. Итак, пишем:

	2	0	-5	0	-8	1
3	2	$3 \cdot 2 + 0 = 6$	$3 \cdot 6 - 5 = 13$	$3 \cdot 13 + 0 = 39$	$3 \cdot 39 - 8 = 109$	$3 \cdot 109 + 1 = 328$

Таким образом, частное равно

$$q(x) = 2x^4 + 6x^3 + 13x^2 + 39x + 109,$$

а остаток равен 328.

Пример 3. Пользуясь схемой Горнера, разделить многочлен

$$f(x) = 3x^4 + 2x^3 - x + 10$$

над кольцом целых чисел на $x + 3$.

Составляем схему Горнера; при этом $a=-3$, так как $x+3=x-(-3)$. Выкладки мы будем здесь проводить в стороне, а в схему будем вписывать только окончательные результаты:

	3	2	0	−1	10
−3	3	−7	21	−64	202

Таким образом, частное равно

$$q(x)=3x^3-7x^2+21x-64,$$

а остаток равен 202.

Схема Горнера выгодна не только для проведения деления многочлена $f(x)$ на $x-a$. Она оказывается весьма удобной и для вычисления значения многочлена при $x=a$. А именно, остаток при делении $f(x)$ на $x-a$ мы можем найти при помощи схемы Горнера, а по теореме 9 этот остаток есть не что иное, как значение многочлена при $x=a$.

Так, в примере 3 было найдено при помощи схемы Горнера, что многочлен

$$f(x)=3x^4+2x^3-x+10$$

при делении на $x+3$ даёт в остатке 202. Мы можем отсюда заключить, что $f(-3)=202$. Приведём ещё один пример.

Пример 4. Пользуясь схемой Горнера, вычислить значение $f(-2)$ многочлена

$$f(x)=x^4-8x^3+24x^2-50x+90$$

над кольцом целых чисел.

Проводим вычисления по схеме Горнера

	1	− 8	24	− 50	90
−2	1	−10	44	−138	366

Мы видим отсюда, что $f(-2)=366$.

Займёмся теперь случаем, когда многочлен $f(x)$ делится на $x-a$ без остатка. Этот случай тесно связан с понятием корня.

Определение. *Корнем многочлена $f(x)$ называется такое значение x_0 неизвестного, при котором значение многочлена равно нулю:* $f(x_0)=0$.

Оказывается, что *элемент a кольца R тогда и только тогда является корнем $f(x)$, когда $f(x)$ делится на $x-a$.*

Доказательство. Если $f(x)$ делится на $x-a$, то по определению делимости должно иметь место равенство

$$f(x)=(x-a)\,q(x).$$

Полагая $x=a$, получаем из этого равенства, что $f(a)=0$, т. е. a оказалось корнем многочлена $f(x)$.

Обратно, пусть a — корень $f(x)$. Тогда по теореме 9 остаток r от деления $f(x)$ на $x-a$ должен равняться $f(a)=0$, т. е. $f(x)$ делится на $x-a$.

Иногда вместо того, чтобы говорить о корне многочлена, говорят о корне алгебраического уравнения n-й степени над кольцом R, т. е. уравнения вида

$$a_0x^n+\ldots+a_n=0,\quad a_0\neq 0, \tag{3}$$

где $a_0, a_1, \ldots, a_n$ — элементы кольца R, называемые коэффициентами уравнения. При этом под корнем уравнения (3) подразумевается корень многочлена $f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n$.

Равенство (3) нельзя, конечно, рассматривать как равенство двух многочленов (многочлена $f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n$ и нуль-многочлена); величина x имеет здесь другой смысл, чем в многочлене: x здесь означает любой из корней рассматриваемого уравнения.

Может случиться, что многочлен $f(x)$ n-й степени будет делиться не только на $x-a$, но и на некоторую степень $x-a$. В соответствии с этим условимся a называть k-кратным корнем многочлена $f(x)$, если $f(x)$ делится на $(x-a)^k$, но не делится на $(x-a)^{k+1}$. Например, если $f(x)$ делится на $(x-a)^2$, но не делится на $(x-a)^3$, то a — двукратный корень $f(x)$ (или корень кратности 2).

Пример 5. Число 1 является корнем многочлена

$$f(x)=x^5-2x^4+x^3+x^2-2x+1$$

над кольцом целых чисел. Найти кратность этого корня.

Делим $f(x)$ на $x-1$ при помощи схемы Горнера:

	1	−2	1	1	−2	1
1	1	−1	0	1	−1	0

Отсюда частное равно

$$q(x)=x^4-x^3+x-1,$$

а остаток, как и следовало ожидать, равен нулю. Получившееся частное делим в свою очередь на $x-1$:

	1	−1	0	1	−1
1	1	0	0	1	0

Мы видим, что и здесь остаток равен нулю, а частное равно

$$q_1(x)=x^3+1.$$

Если теперь разделить $q_1(x)$ на $x-1$, то получится остаток, уже отличный от нуля. Таким образом, данный многочлен $f(x)$ делится на $(x-1)^2$, но не делится на $(x-1)^3$, в силу чего 1 является двукратным корнем $f(x)$.

Возникает естественный вопрос, сколько корней может иметь многочлен $f(x)$ n-й степени над кольцом R. Обратимся к конкретным примерам. Они помогут нам притти к правильному ответу.

Пример 6. Многочлен x^3-2 над полем рациональных чисел не имеет корней. Однако если x^3-2 рассматривать как многочлен над полем действительных чисел, то x^3-2 будет иметь один корень $\sqrt[3]{2}$. Мы видим, что как в случае поля рациональных, так и в случае поля действительных чисел количество корней многочлена x^3-2 меньше трёх, т. е. меньше степени многочлена.

Пример 7. Многочлен x^2-1 над кольцом целых чисел имеет два корня: 1 и −1. Здесь получается, что число корней многочлена равно степени многочлена.

Эти два примера заставляют нас склоняться к мысли, что число корней многочлена n-й степени над R не должно превосходить степени многочлена. Однако следующий пример покажет, что дело обстоит сложнее.

Пример 8. Рассмотрим в качестве кольца R множество квадратных матриц вида

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

где a, b — действительные числа. Предоставляем читателю проверить самому, что это множество в самом деле образует коммутативное кольцо относительно операций сложения и умножения матриц. Единицей здесь, очевидно, будет единичная матрица

$$\varepsilon=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Выясним, какие корни может иметь многочлен $f(x)=x^2-\varepsilon$ над этим кольцом R. Согласно определению корня многочлена нам надо найти такие матрицы:

$$\xi=\begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix},$$

чтобы $\xi^2-\varepsilon=0$, или $\xi^2=\varepsilon$, или в более подробной записи

$$\begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix}^2=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Возводя матрицу в квадрат, получаем:

$$\begin{pmatrix} u^2 & 0 \\ 0 & v^2 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

откуда $u^2=1$, $v^2=1$. Таким образом, $u=\pm 1$, $v=\pm 1$, и мы получаем четыре корня:

$$\xi_1=\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\quad \xi_2=\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},\quad \xi_3=\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},\quad \xi_4=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}=\varepsilon.$$

Как видим, число корней здесь превосходит степень многочлена $f(x)=x^2-\varepsilon$. Вместе с тем рассматриваемое кольцо R содержит делители нуля. Так,

$$\alpha=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\neq 0,\quad \beta=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\neq 0,$$

но

$$\alpha\beta=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}=0.$$

Подведём некоторый итог. В первых двух примерах число корней не превосходило степени многочлена, и в качестве кольца мы имели поле рациональных чисел и кольцо целых чисел — области, не содержащие делители нуля. Напротив, в третьем примере мы имели дело с кольцом, содержащим делители нуля, и число корней рассматриваемого многочлена оказалось больше его степени. Нижеследующая теорема показывает, что эта связь с делителями нуля не случайна.

Теорема 10. *Пусть коммутативное кольцо R с единицей $e\neq 0$ не обладает делителями нуля* (является областью целостности). *Тогда всякий многочлен n-й степени над R имеет в R не более чем n корней, если даже считать каждый корень столько раз, какова его кратность.*

Доказательство. Обозначим через $f(x)$ какой нибудь многочлен над R-выше нулевой степени, и пусть он имеет корни $a_1, \ldots, a_s$

соответственно с кратностями $k_1, \ldots, k_s$. Так как кратность корня a_1 равна k_1, то мы можем написать, что

$$f(x) = (x - a_1)^{k_1} f_1(x),$$

где $f_1(x)$ — многочлен над R, не делящийся на $x - a_1$, т. е. не имеющий a_1 корнем: $f_1(a_1) \neq 0$.

Полагая в этом равенстве $x = a_2$, получаем:

$$(a_2 - a_1)^{k_1} f_1(a_2) = 0.$$

Но $a_2 - a_1 \neq 0$. Следовательно, так как R не имеет делителей нуля, должно быть $f_1(a_2) = 0$. Таким образом, a_2 оказалось корнем многочлена $f_1(x)$.

Обозначим через s кратность корня a_2 относительно $f_1(x)$. Тогда

$$f_1(x) = (x - a_2)^s f_2(x),$$

причём $f_2(a_2) \neq 0$.

Легко видеть, что $s \leqslant k_2$. В самом деле, если бы s было больше k_2, то из равенства

$$f(x) = (x - a_1)^{k_1} (x - a_2)^s f_2(x)$$

следовало бы, что a_2 — корень многочлена $f(x)$ более высокой кратности, чем k_2.

С другой стороны, так как a_2 есть k_2-кратный корень $f(x)$, то

$$f(x) = (x - a_2)^{k_2} \varphi(x),$$

причём $\varphi(a_2) \neq 0$. Отсюда

$$(x - a_1)^{k_1} (x - a_2)^s f_2(x) = (x - a_2)^{k_2} \varphi(x). \qquad (3')$$

Кольцо многочленов $R[x]$ не содержит делителей нуля, так как R — область целостности. Поэтому обе части равенства (3') можно сократить на $(x - a_2)^s$. Получаем:

$$(x - a_1)^{k_1} f_2(x) = (x - a_2)^{k_2 - s} \varphi(x).$$

Если допустить, что $k_2 > s$, то, полагая $x = a_2$, мы имели бы:

$$(a_2 - a_1)^{k_1} f_2(a_2) = 0,$$

откуда $f_2(a_2) = 0$, что невозможно. Следовательно, $k_2 = s$.

Итак,

$$f_1(x) = (x - a_2)^{k_2} f_2(x),$$

а потому

$$f(x) = (x - a_1)^{k_1} (x - a_2)^{k_2} f_2(x).$$

Затем подобным же образом убеждаемся, что

$$f_2(x) = (x - a_3)^{k_3} f_3(x),$$

причём $f_3(a_3) \neq 0$, и

$$f(x) = (x - a_1)^{k_1}(x - a_2)^{k_2}(x - a_3)^{k_3} f_3(x)$$

и т. д. В конечном счёте мы придём к разложению

$$f(x) = (x - a_1)^{k_1}(x - a_2)^{k_2} \ldots (x - a_s)^{k_s} f_s(x).$$

Степень левой части последнего равенства равна n, а степень правой части не меньше чем $k_1 + k_2 + \ldots + k_s$. Отсюда получается, что $k_1 + k_2 + \ldots + k_s \leqslant n$, и теорема для многочлена степени $n \geqslant 1$ доказана. Но для многочлена нулевой степени теорема тривиальна: такой многочлен не имеет корней.

Отметим одно важное следствие.

Следствие. *Если коммутативное кольцо R с единицей $e \neq 0$ является областью целостности и два многочлена $f(x)$, $g(x)$ над R со степенью, не превосходящей n, имеют равные значения более чем при n различных значениях x, то эти многочлены равны:*

$$f(x) = g(x).$$

В самом деле, многочлен $h(x) = f(x) - g(x)$, с одной стороны, имеет степень, не превосходящую n. С другой стороны, $h(x)$ обращается в нуль при более чем n различных значениях x, т. е. имеет более чем n корней. Отсюда по только что доказанной теореме получается, что $h(x) = f(x) - g(x) = 0$, т. е. $f(x) = g(x)$.

В случае бесконечной области целостности R из этого следствия вытекает, что два многочлена $f(x)$ и $g(x)$, имеющие равные значения при любых значениях x, должны быть равны.

Будем в произвольном многочлене $f(x)$ из $R[x]$ неизвестное x заменять тем или иным элементом c кольца R. Мы получим вполне определённый элемент $f(c)$ из R. Таким образом, каждому многочлену $f(x)$ из $R[x]$ будет ставиться в соответствие функция от одного аргумента, определённая на множестве R:

$$f(x) \to f(\xi). \tag{4}$$

Через ξ мы здесь обозначили аргумент, а через $f(\xi)$ — функцию, соответствующую многочлену $f(x)$.

Мы собираемся показать, что в случае бесконечной области целостности R функциональная и алгебраическая точки зрения на многочлен являются равносильными. А именно, имеет место следующая

Теорема 11. *Если коммутативное кольцо R с единицей $e \neq 0$ является бесконечной областью целостности, то множество функций $f(\xi)$, соответствующих многочленам $f(x)$ из $R[x]$, образует кольцо, изоморфное кольцу многочленов $R[x]$.*

Доказательство. Пусть некоторому многочлену $g(x)$ из $R[x]$ соответствует та же функция $f(\xi)$, что и многочлену $f(x)$:

$$f(x) \to f(\xi), \quad g(x) \to f(\xi).$$

Тогда $f(c) = g(c)$ для любого элемента c из R. Но мы уже знаем, что в случае бесконечной области целостности R два многочлена, имеющих равные значения при любых значениях x, должны быть равны. Следовательно, $f(x) = g(x)$. Таким образом, соответствие (4) является не только однозначным, но и взаимно однозначным.

Пусть, далее, $f(x)$ и $g(x)$ — два произвольных многочлена из $R[x]$. Обозначим $f(x) + g(x)$ через $h(x)$ и $f(x)g(x)$ через $k(x)$. Тогда $f(x) + g(x)$ будет соответствовать $h(\xi)$, а $f(x)g(x)$ будет соответствовать $k(\xi)$:

$$f(x) + g(x) \to h(\xi), \quad f(x)g(x) \to k(\xi).$$

Но мы знаем, что для любого элемента c из R

$$f(c) + g(c) = h(c), \quad f(c)g(c) = k(c)$$

(см. равенства (7) в § 1). Следовательно, по определению суммы и произведения функций [1])

$$h(\xi) = f(\xi) + g(\xi), \quad k(\xi) = f(\xi)g(\xi),$$

откуда

$$f(x) + g(x) \to f(\xi) + g(\xi), \quad f(x)g(x) \to f(\xi)g(\xi).$$

Итак, мы убедились, что соответствие (4) в самом деле является изоморфизмом между кольцом $R[x]$ и множеством функций $f(\xi)$. Тем самым множество функций $f(\xi)$ образует кольцо, изоморфное $R[x]$, и теорема полностью доказана.

В дальнейшем мы будем аргумент функций $f(\xi)$ обозначать той же буквой x, что и неизвестное.

§ 4. Многочлены над полем рациональных чисел

В элементарной алгебре рассматриваются простейшие методы разложения многочлена $f(x)$ с рациональными коэффициентами на неприводимые множители в поле рациональных чисел. Так как эти методы связаны с вычислением рациональных корней многочленов, то мы в этом параграфе изложим с необходимой полнотой вопрос о вычислении рациональных корней.

---

[1]) Пусть M — некоторое множество с двумя определёнными в нём алгебраическими операциями $+$ и $\cdot$. Под суммой $f(\xi) + g(\xi)$ двух функций $f(\xi)$ и $g(\xi)$, заданных на множестве M, подразумевается функция, ставящая в соответствие каждому элементу c из M сумму значений $f(c) + g(c)$ данных функций при $\xi = c$. Точно так же под произведением $f(\xi)g(\xi)$ функций $f(\xi)$ и $g(\xi)$ подразумевается функция, ставящая в соответствие каждому элементу c из M произведение значений $f(c)g(c)$ данных функций при $\xi = c$.

Итак, пусть

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n \qquad (a_0\neq 0) \tag{1}$$

— многочлен n-й степени ($n\geqslant 1$) с рациональными коэффициентами. Без ограничения общности выводов можно предположить, что все коэффициенты многочлена (1) — числа целые. Действительно, если бы многочлен $f(x)$ имел дробные коэффициенты, то, умножая $f(x)$ на общий знаменатель коэффициентов, мы получили бы многочлен с целыми коэффициентами и с теми же корнями, что и у $f(x)$.

Вычисление рациональных корней многочлена (1) основано на следующей теореме.

Теорема 12. *Если несократимая дробь* $\frac{l}{m}$ (l, m — целые числа) *является рациональным корнем многочлена* (1), *то* l *есть делитель свободного члена* a_n, *а* m *— делитель старшего коэффициента* a_0.

Доказательство. Согласно определению корня многочлена мы можем написать, что

$$a_0\frac{l^n}{m^n}+a_1\frac{l^{n-1}}{m^{n-1}}+\ldots+a_{n-1}\frac{l}{m}+a_n=0,$$

или, умножая обе части последнего равенства на m^n:

$$a_0l^n+a_1l^{n-1}m+\ldots+a_{n-1}lm^{n-1}+a_nm^n=0.$$

Отсюда

$$a_0l^n=-m(a_1l^{n-1}+\ldots+a_nm^{n-1}) \tag{2}$$

и

$$a_nm^n=-l(a_0l^{n-1}+\ldots+a_{n-1}m^{n-1}). \tag{3}$$

Правая часть равенства (2) делится, очевидно, на m. Следовательно, на m должна делиться и левая часть равенства (2), т. е. a_0l^n. Но в силу несократимости дроби $\frac{l}{m}$ число l^n взаимно просто с m. Поэтому a_0 должно делиться на m.

Аналогично рассуждаем и относительно равенства (3). Его правая часть делится на l. Следовательно, a_nm^n должно также делиться на l. Отсюда a_n делится на l, так как m^n взаимно просто с l.

Отметим одно следствие из только что доказанной теоремы.

Следствие. *Многочлен*

$$f(x)=x^n+a_1x^{n-1}+\ldots+a_n$$

со старшим коэффициентом, равным единице, и с целыми коэффициентами $a_1, \ldots, a_n$ *может иметь в качестве рациональных корней только целые корни.*

В самом деле, по теореме 12 знаменатель $m>0$ рационального корня $x_0=\frac{l}{m}$ должен быть делителем старшего коэффициента,

т. е. равен 1. Отсюда $x_0 = l$, и тем самым корень x_0 является целым числом.

Таким образом, испытывая всевозможные дроби $\frac{l}{m}$ $(m > 0)$ с числителем l, делящим a_n, и со знаменателем m, делящим старший коэффициент a_0, мы найдём рациональные корни многочлена (1) или убедимся, что многочлен (1) вовсе не имеет рациональных корней. Однако эти испытания можно значительно сократить, если воспользоваться следующим предложением.

Теорема 13. *Если несократимая дробь* $\frac{l}{m}$ $(m > 0)$ *является рациональным корнем многочлена* (1), *то для любого целого числа* k *число* $f(k)$ *делится на* $l - km$ *при условии, что* $l - km \neq 0$.

Доказательство. Умножая многочлен (1) на m^n, получаем:

$$m^n f(x) = a_0 (mx)^n + m a_1 (mx)^{n-1} + \ldots + m^n a_n,$$

или, полагая $mx = y$:

$$m^n f(x) = \varphi(y) = a_0 y^n + m a_1 y^{n-1} + \ldots + m^n a_n.$$

Так как $\frac{l}{m}$ — корень многочлена $f(x)$, то целое число l должно быть корнем многочлена $\varphi(y)$, в силу чего мы можем написать, рассматривая $\varphi(y)$ над кольцом целых чисел, что

$$\varphi(y) = (y - l)\, q(y),$$

где $q(y)$ есть также многочлен над кольцом целых чисел. Отсюда

$$\frac{\varphi(km)}{l - km} = \frac{m^n f(k)}{l - km} = -q(km) \tag{4}$$

должно быть целым числом; иными словами, $m^n f(k)$ делится на $l - km$. Но легко видеть, что m и $l - km$ взаимно просты. В самом деле, если бы m и $l - km$ были бы не взаимно простыми, то дробь

$$\frac{l - km}{m} = \frac{l}{m} - k$$

была бы сократимой:

$$\frac{l - km}{m} = \frac{l_1}{m_1},$$

где $0 < m_1 < m$, и мы имели бы, что

$$\frac{l_1}{m_1} = \frac{l}{m} - k,$$

откуда

$$\frac{l}{m} = \frac{l_1 + km_1}{m_1},$$

т. е. в силу неравенства $m_1 < m$ следовала бы сократимость дроби $\frac{l}{m}$, что невозможно.

Теперь теорема становится очевидной — произведение $m^n f(k)$ делится на $l-km$, а m взаимно просто с $l-km$; следовательно, $f(k)$ делится на $l-km$.

Обращаясь к примерам, покажем, как на основании изложенного следует вычислять рациональные корни многочлена.

Пример 1. Найти рациональные корни многочлена

$$f(x)=x^4-2x^3-8x^2+13x-24.$$

В этом многочлене все коэффициенты являются целыми числами и старший коэффициент равен единице. Следовательно, если многочлен $f(x)$ имеет рациональные корни, то согласно следствию из теоремы 12 корни должны быть целыми. На основании теоремы 12 заключаем, что целые корни многочлена $f(x)$ должны быть делителями его свободного члена — 24. Таким образом, целые корни следует искать среди чисел

$$l=1,\ -1,\ 2,\ -2,\ 3,\ -3,\ 4,\ -4,\ 6,\ -6,\ 8,\ -8,\ 12,\ -12,\ 24,\ -24. \quad (5)$$

Эти числа можно рассматривать как дроби $\frac{l}{m}$ с $m=1$; отсюда на основании теоремы 13 заключаем, что для целого корня $x_0=\frac{l}{1}=l$ число $f(k)$ должно делиться на $l-k$, где k — произвольное целое число, отличное от l. Возьмём $k=1$ и $k=-1$. Так как $f(1)=-20$ и $f(-1)=-42$, то 1 и -1 не могут быть корнями рассматриваемого многочлена, и потому остаётся исследовать числа

$$l=2,\ -2,\ 3,\ -3,\ 4,\ -4,\ 6,\ -6,\ 8,\ -8,\ 12,\ -12,\ 24,\ -24. \quad (6)$$

Посмотрим, для каких чисел (6) $f(1)$ делится на $l-1$ и $f(-1)$ делится на $l+1$. Легко усмотреть, что только числа

$$l=2,\ -3,\ -4,\ 6$$

удовлетворяют этому условию.

Так как $f(2)=-30\neq 0$, то остаётся исследовать

$$l=-3,\ -4 \text{ и } 6.$$

Число $f(2)=-30$ должно делиться на $l-2$, если l — целый корень многочлена $f(x)$. Но этому условию делимости удовлетворяют только -3 и -4. Подставляя в выражение многочлена вместо x значения -3 и -4, видим, что $f(-3)=0$ и $f(-4)=180$. Таким образом, рассматриваемый многочлен $f(x)$ имеет только один рациональный корень $x_0=-3$.

Пример 2. Найти рациональные корни многочлена

$$f(x)=24x^5+10x^4-x^3-19x^2-5x+6.$$

По теореме 12 знаменатель m рационального корня $x_0 = \frac{l}{m}$ многочлена $f(x)$ должен быть делителем 24, а числитель l — делителем 6. Мы можем знаменатель m считать положительным, относя знак к числителю l. В соответствии со всем этим составляем следующую таблицу возможных значений x_0:

l	m	x_0	l	m	x_0	l	m	x_0
1	1	1	3	2	$3/2$	1	6	$1/6$
−1	1	−1	−3	2	$-3/2$	−1	6	$-1/6$
2	1	2	1	3	$1/3$	1	8	$1/8$
−2	1	−2	−1	3	$-1/3$	−1	8	$-1/8$
3	1	3	2	3	$2/3$	3	8	$3/8$
−3	1	−3	−2	3	$-2/3$	−3	8	$-3/8$
6	1	6	1	4	$1/4$	1	12	$1/12$
−6	1	−6	−1	4	$-1/4$	−1	12	$-1/12$
1	2	$1/2$	3	4	$3/4$	1	24	$1/24$
−1	2	$-1/2$	−3	4	$-3/4$	−1	24	$-1/24$

Числа 1 и −1 не могут быть корнями многочлена $f(x)$, так как $f(1)=15$ и $f(-1)=-21$. Затем, пользуясь теоремой 13, можно исключить ещё ряд возможных значений x_0. А именно, испытываем, для каких $\frac{l}{m}$ число $f(1)=15$ делится на $l-m$ и число $f(-1)=-21$ делится на $l+m$. Без труда находим, что этим условиям делимости удовлетворяют только числа

$$x_0 = 2,\ -2,\ 6,\ 1/2,\ -1/2,\ -3/2,\ -2/3,\ -1/4,\ 3/4,\ 1/6.$$

Так как $f(2)=840$ и $f(-2)=-660$, то 2 и −2 исключаются и остаётся исследовать числа

$$6,\ 1/2,\ -1/2,\ -3/2,\ -2/3,\ -1/4,\ 3/4,\ 1/6. \tag{7}$$

Выкидываем из ряда (7) те числа $\frac{l}{m}$, для которых $f(2)=840$ не делится на $l-2m$. Затем выкидываем $\frac{l}{m}$, для которых $f(-2)=-660$ не делится на $l+2m$. В результате останутся

$$1/2,\ -1/2,\ -3/2,\ -2/3,\ 3/4.$$

Подвергаем эти значения непосредственному испытанию — подставляем вместо x эти значения в многочлен $f(x)$:

$$f\left(\frac{1}{2}\right)=f\left(-\frac{2}{3}\right)=f\left(\frac{3}{4}\right)=0,\ f\left(-\frac{1}{2}\right)=\frac{15}{4},\ f\left(-\frac{3}{2}\right)=-162.$$

Мы видим отсюда, что рассматриваемый многочлен имеет только следующие рациональные корни:

$$x_0 = \frac{1}{2},\ -\frac{2}{3},\ \frac{3}{4},$$

и мы можем написать разложение $f(x)$ в произведение неприводимых (в поле рациональных чисел) многочленов

$$f(x) = \left(x - \frac{1}{2}\right)\left(x + \frac{2}{3}\right)\left(x - \frac{3}{4}\right)(24x^2 + 24x + 24)$$

или

$$f(x) = (2x - 1)(3x + 2)(4x - 3)(x^2 + x + 1).$$

Квадратный трёхчлен $x^2 + x + 1$ неприводим, так как он не имеет рациональных корней.

В некоторых случаях большую помощь при вычислении рациональных корней могут оказать следующие предложения:

1°. *Многочлен $f(x)$ с целыми коэффициентами не имеет целых корней, если для некоторых целых s и t $f(2s)$ и $f(2t+1)$ — нечётные числа.*

Доказательство. Допустим противное — предположим, что $f(x)$ имеет целый корень x_0. Тогда

$$f(x) = (x - x_0)\,q(x).$$

Отсюда

$$f(2s) = (2s - x_0)\,q(2s) \tag{8}$$

и

$$f(2t+1) = (2t + 1 - x_0)\,q(2t+1). \tag{9}$$

Из равенства (8) следует в силу нечётности $f(2s)$, что число $2s - x_0$ должно быть нечётным. Так как $2s$ является чётным, то отсюда получается, что x_0 должно быть нечётным числом. С другой стороны, из равенства (9) следует, что $2t + 1 - x_0$ должно быть нечётным, откуда благодаря нечётности $2t + 1$ вытекает, что x_0 должно быть чётным. Итак, одно и то же число x_0 оказывается одновременно чётным и нечётным, что абсурдно.

2°. *Многочлен $f(x)$ с целыми коэффициентами не имеет рациональных корней, если можно указать два таких целых значения k_1 и k_2 независимого переменного x, что $k_1 - k_2 > 2$ и $f(k_1) = \pm 1$ и $f(k_2) = \pm 1$.*

Доказательство. Допустим, что многочлен $f(x)$ при указанных условиях имеет рациональный корень $x_0 = \frac{l}{m}$. Тогда по теореме 13 $f(k_1) = \pm 1$ должно делиться на $l - k_1 m$ и $f(k_2) = \pm 1$ должно делиться на $l - k_2 m$. Отсюда получается, что

$$l - k_1 m = \pm 1 \quad \text{и} \quad l - k_2 m = \pm 1.$$

Вычитая из второго равенства первое, мы будем иметь:

$$(k_1 - k_2)\, m = \pm 2 \quad \text{или} \quad (k_1 - k_2)\, m = 0.$$

Но равенство $(k_1 - k_2)\, m = 0$ исключается, так как $k_1 \neq k_2$ и $m > 0$. Таким образом, имеет место только первое равенство. Из этого равенства видно, что 2 должно делиться на $k_1 - k_2$, что невозможно, так как по условию $k_1 - k_2 > 2$.

Пример 3. Найти рациональные корни многочлена

$$f(x) = x^6 + x^5 - x^4 - 2x^3 - 6x^2 + 7x + 105.$$

Согласно следствию из теоремы 12 этот многочлен может иметь в качестве рациональных корней лишь целые корни. Легко видеть, что $f(0)$ и $f(1)$ — здесь нечётные числа: $f(0) = f(1) = 105$. Следовательно, согласно предложению 1° рассматриваемый многочлен не имеет целых корней и тем самым не имеет рациональных корней.

Пример 4. Найти рациональные корни многочлена

$$f(x) = 2x^4 - 7x^3 - x^2 - 18x + 25.$$

Полагая $x = 1$ и $x = 4$, получаем: $f(1) = 1$ и $f(4) = 1$. Мы видим, что для этого многочлена условия предложения 2° соблюдаются. Следовательно, рассматриваемый многочлен не имеет рациональных корней.

§ 5. Разложение многочленов на неприводимые множители над полем рациональных чисел. Признак неприводимости

На основании изложенного в предыдущем параграфе мы теперь укажем различные приёмы разложения многочленов $f(x)$ с рациональными коэффициентами на множители, неприводимые в поле рациональных чисел.

Для многочленов второй и третьей степени вопрос о разложении на множители решается просто. А именно, если многочлен второй степени

$$f(x) = ax^2 + bx + c \tag{1}$$

с рациональными коэффициентами приводим в поле рациональных чисел, то он, очевидно, будет разлагаться в произведение двух линейных множителей:

$$f(x) = a(x - x_1)(x - x_2)$$

(x_1 и x_2 — рациональные числа), вследствие чего $f(x)$ будет иметь два рациональных корня x_1 и x_2. Обратно, если многочлен (1) имеет хотя бы один рациональный корень x_1, то многочлен будет разлагаться в поле рациональных чисел в произведение двух линейных множителей.

Примерно такую же роль играют рациональные корни и для многочлена третьей степени

$$f(x)=ax^3+bx^2+cx+d \tag{2}$$

с рациональными коэффициентами: *многочлен* (2) *приводим в поле рациональных чисел тогда и только тогда, когда он имеет по меньшей мере один рациональный корень.*

В самом деле, если многочлен (2) приводим в поле рациональных чисел, то он должен иметь в своём разложении хотя бы один линейный множитель $px+q$ с рациональными коэффициентами p, q. Этот множитель обладает рациональным корнем $x_0=-\frac{q}{p}$. Очевидно, что $x_0=-\frac{q}{p}$ будет также корнем и многочлена (2).

Обратно, если многочлен (2) имеет рациональный корень x_0, то

$$f(x)=(x-x_0)\,q(x),$$

откуда $f(x)$ приводим в поле рациональных чисел.

Пример 1. Разложить многочлен

$$f(x)=2x^3-x^2-x-1$$

на множители в поле рациональных чисел.

Пользуясь методами предыдущего параграфа, находим, что данный многочлен не имеет рациональных корней. Следовательно, этот многочлен неприводим в поле рациональных чисел.

Пример 2. Разложить многочлен

$$f(x)=6x^3-7x^2-2x+2$$

на множители в поле рациональных чисел.

Находим, что этот многочлен имеет только один рациональный корень, равный $\frac{1}{2}$. Следовательно, данный многочлен приводим в поле рациональных чисел, а именно, распадается на линейный и квадратный множители. С помощью схемы Горнера без труда находим искомое разложение на множители:

$$f(x)=\left(x-\frac{1}{2}\right)(6x^2-4x-4)$$

или

$$f(x)=(2x-1)(3x^2-2x-2).$$

Пример 3. Разложить многочлен

$$f(x)=6x^3+7x^2-9x+2$$

на множители в поле рациональных чисел.

Находим, что многочлен $f(x)$ имеет три рациональных корня $\frac{1}{2}$, $\frac{1}{3}$ и -2. Следовательно, $f(x)$ должен разлагаться на три линейных множителя $\left(x-\frac{1}{2}\right)$, $\left(x-\frac{1}{3}\right)$ и $(x+2)$:

$$f(x)=6\left(x-\frac{1}{2}\right)\left(x-\frac{1}{3}\right)(x+2).$$

Впереди стоит число 6, так как старший коэффициент $f(x)$ равен 6. Освобождаясь от знаменателей, получаем окончательно, что

$$f(x)=(2x-1)(3x-1)(x+2).$$

Для многочленов более высокой степени дело обстоит сложнее — если многочлен n-й степени ($n\geqslant 4$) с рациональными коэффициентами имеет рациональный корень x_0, то этот многочлен будет приводим в поле рациональных чисел, так как будет делиться на $x-x_0$. Обратное, однако, неверно. Например, многочлен

$$f(x)=x^5-x^4-3x^3+x^2+3x+1$$

не имеет рациональных корней и тем не менее он приводим в поле рациональных чисел:

$$f(x)=(x^2-x-1)(x^3-2x-1).$$

Для многочленов четвёртой степени можно указать довольно удобный приём разложения на множители, связанный с понятием кубической резольвенты.

Пусть

$$f(x)=x^4+ax^3+bx^2+cx+d \tag{3}$$

— некоторый многочлен четвёртой степени с рациональными коэффициентами. Преобразуем его так, чтобы он представился в виде разности двух квадратов. Для этой цели пишем:

$$f(x)=\left[(x^2)^2+2x^2\left(\frac{ax}{2}\right)\right]+(bx^2+cx+d).$$

Сделаем выражение внутри квадратных скобок полным квадратом, для чего прибавим и вычтем $\left(\frac{ax}{2}\right)^2$:

$$f(x)=\left[(x^2)^2+2x^2\cdot\left(\frac{ax}{2}\right)+\left(\frac{ax}{2}\right)^2\right]+\left[\left(b-\frac{a^2}{4}\right)x^2+cx+d\right],$$

$$f(x)=\left(x^2+\frac{ax}{2}\right)^2+\left[\left(b-\frac{a^2}{4}\right)x^2+cx+d\right].$$

Далее, введём вспомогательную величину y, а именно прибавим к последнему выражению и вычтем из него многочлен

$$2\left(x^2+\frac{ax}{2}\right)y+y^2.$$

Получим:

$$f(x)=\left(x^2+\frac{ax}{2}+y\right)^2+$$
$$+\left[\left(b-\frac{a^2}{4}\right)x^2+cx+d-2\left(x^2+\frac{ax}{2}\right)y-y^2\right]$$

или

$$f(x)=\left(x^2+\frac{ax}{2}+y\right)^2-(Ax^2+Bx+C), \qquad (4)$$

где

$$A=2y+\frac{a^2}{4}-b, \quad B=ay-c, \quad C=y^2-d.$$

Теперь подберём y таким, чтобы квадратный трёхчлен Ax^2+Bx+C был полным квадратом. Мы воспользуемся следующим предложением:

Квадратный трёхчлен Ax^2+Bx+C с комплексными коэффициентами A, B, C тогда и только тогда является квадратом линейного многочлена $\alpha x+\beta$ с комплексными коэффициентами, когда $B^2=4AC$.

Доказательство. Пусть

$$Ax^2+Bx+C=(\alpha x+\beta)^2.$$

Тогда

$$Ax^2+Bx+C=\alpha^2x^2+2\alpha\beta x+\beta^2.$$

Мы знаем, что если два многочлена равны, то у них должны совпадать коэффициенты при одинаковых степенях x. Следовательно,

$$A=\alpha^2, \quad B=2\alpha\beta, \quad C=\beta^2.$$

Так как, очевидно, $(2\alpha\beta)^2=4\alpha^2\beta^2$, то отсюда вытекает, что $B^2=4AC$.

Обратно, пусть $B^2=4AC$. Тогда квадратный трёхчлен можно преобразовать следующим образом:

$$Ax^2+Bx+C=(\sqrt{A}\,x)^2+2(\sqrt{A}\,x)\sqrt{C}+(\sqrt{C})^2=$$
$$=(\sqrt{A}\,x+\sqrt{C})^2,$$

т. е. Ax^2+Bx+C представился в виде квадрата линейного двучлена.

Вернёмся теперь к выражению (4) многочлена $f(x)$. На основании только что доказанного предложения попытаемся подобрать y с таким расчётом, чтобы $B^2=4AC$ или

$$(ay-c)^2=4\left(2y+\frac{a^2}{4}-b\right)(y^2-d). \qquad (5)$$

Мы получили уравнение третьей степени относительно y. Это уравнение (5) и называется *кубической резольвентой* многочлена (3).

Итак, *если y — один из корней кубической резольвенты* (5), *то многочлен* (3) *будет выражаться в виде разности двух квадратов*:

$$f(x)=\left(x^2+\frac{ax}{2}+y\right)^2-(\alpha x+\beta)^2.$$

Предлагаемый вниманию читателей приём разложения многочлена (3) четвёртой степени на множители в поле рациональных чисел основан на следующей теореме:

Теорема 14. *Многочлен четвёртой степени*

$$f(x)=x^4+ax^3+bx^2+cx+d \qquad (3')$$

с рациональными коэффициентами, не имеющий рациональных корней, тогда и только тогда приводим в поле рациональных чисел, когда его резольвента обладает таким рациональным корнем y_0, что

$$\sqrt{2y_0+\frac{a^2}{4}-b}, \quad \text{и} \quad \sqrt{y_0^2-d}$$

являются рациональными числами.

Доказательство. Пусть резольвента (5) имеет рациональный корень y_0 и $\sqrt{2y_0+\frac{a^2}{4}-b}$, $\sqrt{y_0^2-d}$ — рациональные числа. Тогда на основании вышеизложенного мы можем написать:

$$f(x)=\left(x^2+\frac{ax}{2}+y_0\right)^2-(\alpha x+\beta)^2,$$

где

$$\alpha=\pm\sqrt{2y_0+\frac{a^2}{4}-b}, \quad \beta=\pm\sqrt{y_0^2-d}$$

— рациональные числа. Но разность двух квадратов можно, как известно, представить в виде произведения суммы на разность. Следовательно,

$$f(x)=\left[x^2+\left(\frac{a}{2}+\alpha\right)x+(y_0+\beta)\right]\cdot\left[x^2+\left(\frac{a}{2}-\alpha\right)x+(y_0-\beta)\right],$$

т. е. многочлен $f(x)$ оказался приводимым в поле рациональных чисел.

Обратно, пусть $f(x)$ приводим в поле рациональных чисел. Так как по условию $f(x)$ не имеет рациональных корней, то многочлен $f(x)$ будет разлагаться в произведение двух квадратных трёхчленов с рациональными коэффициентами:

$$f(x)=x^4+ax^3+bx^2+cx+d=(x^2+p_1x+q_1)(x^2+p_2x+q_2)$$

или

$$x^4+ax^3+bx^2+cx+d=$$

$$=x^4+(p_1+p_2)x^3+(p_1p_2+q_1+q_2)x^2+(p_1q_2+p_2q_1)x+q_1q_2.$$

Сравнивая слева и справа коэффициенты при одинаковых степенях x, получаем:

$$p_1+p_2=a,\quad p_1p_2+q_1+q_2=b,\quad p_1q_2+p_2q_1=c,\quad q_1q_2=d. \tag{6}$$

Исходя из равенств (6), нетрудно убедиться, что резольвента (5) имеет рациональным корнем

$$y_0=\frac{1}{2}(q_1+q_2)$$

и $\sqrt{2y_0+\frac{a^2}{4}-b}$, $\sqrt{y_0^2-d}$ являются рациональными числами. В самом деле,

$$A=2y_0+\frac{a^2}{4}-b=(q_1+q_2)+\frac{(p_1+p_2)^2}{4}-(p_1p_2+q_1+q_2)=$$

$$=\frac{p_1^2-2p_1p_2+p_2^2}{4}=\left(\frac{p_1-p_2}{2}\right)^2,$$

$$C=y_0^2-d=\left(\frac{q_1+q_2}{2}\right)^2-q_1q_2=\frac{q_1^2-2q_1q_2+q_2^2}{4}=\left(\frac{q_1-q_2}{2}\right)^2,$$

$$B=ay_0-c=(p_1+p_2)\cdot\frac{q_1+q_2}{2}-(p_1q_2+p_2q_1)=$$

$$=\frac{p_1q_1+p_2q_2-p_1q_2-p_2q_1}{2}=\frac{(p_1-p_2)(q_1-q_2)}{2},$$

откуда уравнение (5) при $y=y_0=\frac{q_1+q_2}{2}$ превращается в очевидное тождество

$$\left[\frac{(p_1-p_2)(q_1-q_2)}{2}\right]^2=4\left(\frac{p_1-p_2}{2}\right)^2\left(\frac{q_1-q_2}{2}\right)^2.$$

Итак, если многочлен (3) приводим в поле рациональных чисел, то его резольвента (5) имеет рациональный корень

$$y_0=\frac{q_1+q_2}{2}$$

и

$$\sqrt{2y_0+\frac{a^2}{4}-b}=\sqrt{A}=\frac{p_1-p_2}{2},\quad \sqrt{y_0^2-d}=\sqrt{C}=\frac{q_1-q_2}{2}$$

— рациональные числа, так как $\frac{p_1-p_2}{2}$ и $\frac{q_1-q_2}{2}$ рациональны.

Покажем теперь на конкретных примерах, как на основании изложенного проводится разложение многочленов четвёртой степени на множители.

Пример 4. Разложить многочлен

$$f(x)=2x^4+x^3-3x^2+3x-1$$

на множители в поле рациональных чисел.

Прежде всего выясним, имеет ли данный многочлен рациональные корни. С помощью обычных приёмов вычисления рациональных корней убеждаемся в том, что $f(x)$ имеет только один рациональный корень $x=\frac{1}{2}$. Пользуясь схемой Горнера, получаем:

$$f(x)=\left(x-\frac{1}{2}\right)(2x^3+2x^2-2x+2)$$

или

$$f(x)=(2x-1)(x^3+x^2-x+1).$$

Многочлен x^3+x^2-x+1 уже неприводим в поле рациональных чисел, так как он не имеет рациональных корней (корни этого многочлена являются также корнями многочлена $f(x)$).

Пример 5. Разложить многочлен

$$f(x)=x^4+3x^3-2x^2+2x-2$$

на множители в поле рациональных чисел.

Этот многочлен не имеет рациональных корней. Поэтому воспользуемся теоремой 14. Составляем резольвенту многочлена $f(x)$:

$$(3y-2)^2=4\left(2y+\frac{9}{4}+2\right)(y^2+2),$$

или после очевидных упрощений

$$8y^3+8y^2+28y+30=0,$$

или, наконец,

$$z^3+2z^2+14z+30=0,$$

где $z=2y$. Последнее уравнение не имеет, однако, рациональных корней. Следовательно, резольвента также не имеет рациональных корней. Отсюда рассматриваемый многочлен неприводим в поле рациональных чисел.

Пример 6. Разложить многочлен

$$f(x)=x^4+2x^3-2x^2+2x+1$$

на множители в поле рациональных чисел.

Нетрудно убедиться, что этот многочлен не имеет рациональных корней. Поэтому обращаемся к теореме 14. Составляем резольвенту многочлена

$$(2y-2)^2=4(2y+3)(y^2-1)$$

или

$$(y-1)(y^2+2y+2)=0.$$

Отсюда легко усмотреть, что резольвента имеет только один рациональный корень $y_0=1$. Для этого корня

$$\sqrt{y_0^2-d}=\sqrt{1-1}=0,\quad \sqrt{2y_0+\frac{a^2}{4}-b}=\sqrt{5}.$$

Итак, рассматриваемый многочлен неприводим в поле рациональных чисел, так как число $\sqrt{5}$ иррационально.

Пример 7. Разложить многочлен

$$f(x)=6x^4-7x^3+x^2-2$$

на множители в поле рациональных чисел.

Этот многочлен не имеет рациональных корней. Чтобы можно было воспользоваться теоремой 14, преобразуем $f(x)$ в многочлен со старшим коэффициентом, равным единице, для чего делим $f(x)$ на 6:

$$f_1(x)=x^4-\frac{7}{6}x^3+\frac{1}{6}x^2-\frac{1}{3}.$$

Составляем резольвенту этого многочлена:

$$\left(-\frac{7}{6}y\right)^2=4\left(2y+\frac{49}{144}-\frac{1}{6}\right)\left(y^2+\frac{1}{3}\right),$$

или окончательно

$$108z^3-18z^2+144z+25=0,$$

где $z=2y$.

Это уравнение имеет рациональный корень $z_0=-\frac{1}{6}$. Отсюда резольвента также имеет рациональный корень, именно $y_0=-\frac{1}{12}$. Для этого корня y_0

$$\sqrt{2y_0+\frac{a^2}{4}-b}=\sqrt{-\frac{1}{6}+\frac{49}{144}-\frac{1}{6}}=\sqrt{\frac{1}{144}}=\frac{1}{12},$$

$$\sqrt{y_0^2-d}=\sqrt{\frac{1}{144}+\frac{1}{3}}=\sqrt{\frac{49}{144}}=\frac{7}{12}.$$

Следовательно, многочлен $f_1(x)$ и тем самым многочлен $f(x)$ приводимы в поле рациональных чисел. Найдём, на какие множители распадается $f(x)$. Для этой цели вычислим B:

$$B=ay_0-c=-\frac{7}{6}\cdot\left(-\frac{1}{12}\right)=\frac{7}{72}.$$

Число B оказалось положительным. Отсюда α и β должны быть одного знака, так как $2\alpha\beta=B>0$. Возьмём α и β со знаком плюс (с таким же успехом можно было их взять и со знаком минус):

$$\alpha=+\sqrt{2y_0+\frac{a^2}{4}-b}=\frac{1}{12},\quad \beta=+\sqrt{y_0^2-d}=\frac{7}{12}.$$

Тогда, руководствуясь доказательством первой половины теоремы 14, получаем:

$$f_1(x)=\left[x^2+\left(\frac{a}{2}+\alpha\right)x+(y_0+\beta)\right]\left[x^2+\left(\frac{a}{2}-\alpha\right)x+(y_0-\beta)\right]=$$
$$=\left[x^2+\left(-\frac{7}{12}+\frac{1}{12}\right)x+\left(-\frac{1}{12}+\frac{7}{12}\right)\right]\times$$
$$\times\left[x^2+\left(-\frac{7}{12}-\frac{1}{12}\right)x+\left(-\frac{1}{12}-\frac{7}{12}\right)\right]=$$
$$=\left(x^2-\frac{1}{2}x+\frac{1}{2}\right)\left(x^2-\frac{2}{3}x-\frac{2}{3}\right).$$

Отсюда

$$f(x)=6f_1(x)=(2x^2-x+1)(3x^2-2x-2).$$

Обращаясь к многочленам $f(x)$ выше четвёртой степени, отметим, прежде всего, что следует всегда начинать с вычисления рациональных корней многочлена. Если $f(x)$ имеет хотя бы один рациональный корень x_0, то $f(x)=(x-x_0)f_1(x)$, и тогда дело сводится к разложению на дальнейшие множители многочлена $f_1(x)$, имеющего меньшую степень. Если же $f(x)$ не имеет рациональных корней, то приходится прибегать к особым методам. Один из таких методов мы и собираемся изложить в общих чертах.

Коэффициенты многочлена

$$f(x)=a_0+a_1x+\ldots+a_nx^n \qquad (a_n\neq 0)$$

над полем рациональных чисел всегда можно предполагать целыми, так как в противном случае мы умножили бы $f(x)$ на общего знаменателя его коэффициентов.

В этом предположении назовём многочлен $f(x)$ *примитивным*, если наибольший общий делитель всех его коэффициентов равен единице.

Например,

$$f(x)=2x^5-4x^4+3x^3+2x^2-5x+7$$

является примитивным многочленом; напротив

$$f(x)=3x^4-21x^3+15x^2-6x+18$$

не примитивен, так как наибольший общий делитель всех его коэффициентов равен 3, а не единице.

Рассмотрим две леммы, известные под названием лемм Гаусса.

Лемма 1. *Произведение двух примитивных многочленов есть также примитивный многочлен.*

Доказательство. Пусть

$$\varphi(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n,$$
$$\psi(x)=b_0+b_1x+b_2x^2+\ldots+b_mx^m$$

— примитивные многочлены. Предположим, вопреки утверждению леммы, что их произведение

$$\varphi(x)\psi(x) = c_0 + c_1x + \ldots + c_{n+m}x^{n+m}$$

не примитивно. Тогда все коэффициенты

$$\begin{aligned}
c_0 &= a_0b_0,\\
c_1 &= a_0b_1 + a_1b_0,\\
&\ldots\ldots\ldots\\
c_{i+j} &= a_0b_{i+j} + a_1b_{i+j-1} + \ldots + a_{i-1}b_{j+1} + a_ib_j +\\
&\qquad + a_{i+1}b_{j-1} + \ldots + a_{i+j}b_0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
c_{n+m} &= a_nb_m
\end{aligned}$$

произведения $\varphi(x)\psi(x)$ будут обладать наибольшим общим делителем d, отличным от единицы. Пусть p — простое число, делящее d. Тогда, очевидно, p будет делить все коэффициенты c_0, c_1, c_2, ..., c_{n+m} произведения $\varphi(x)\psi(x)$. Однако, p не может делить все коэффициенты $\varphi(x)$, так как в противном случае $\varphi(x)$ не было бы примитивным. Точно так же p не может делить все коэффициенты $\psi(x)$. Таким образом, пусть a_i — первый коэффициент многочлена $\varphi(x)$, не делящийся на p, и b_j — первый коэффициент $\psi(x)$, не делящийся на p. Рассмотрим коэффициент

$$c_{i+j} = a_0b_{i+j} + \ldots + a_{i-1}b_{j+1} + a_ib_j + a_{i+1}b_{j-1} + \ldots + a_{i+j}b_0. \quad (6)$$

В правой части равенства (6) все члены, кроме a_ib_j, делятся на p, так как a_0, ..., a_{i-1}, b_{j-1}, ..., b_0 ещё делятся на p. Но член a_ib_j на p не делится, так как a_i и b_j не делятся на p. Отсюда следует, что правая часть равенства (6) не делится на p. Получилось противоречие с нашим предположением о том, что все коэффициенты произведения $\varphi(x)\psi(x)$, и в частности c_{i+j}, делятся на p. Это противоречие и доказывает лемму.

С помощью леммы 1 докажем вторую лемму, играющую в излагаемом методе основную роль. Эта лемма состоит в следующем:

Лемма 2. *Если многочлен $f(x)$ с целыми коэффициентами приводим в поле рациональных чисел, то он разлагается в произведение двух многочленов низшей степени с целыми коэффициентами.*

Доказательство. Пусть многочлен

$$f(x) = a_0 + a_1x + \ldots + a_nx^n \qquad (a_n \neq 0,\ n \geqslant 2)$$

разлагается в поле рациональных чисел следующим образом на два множителя низшей степени:

$$f(x) = g(x)\,h(x).$$

Если все коэффициенты многочленов $g(x)$ и $h(x)$ — числа целые, то доказывать нечего. Поэтому пусть $g(x)$ и $h(x)$ имеют дробные коэффициенты. Обозначим через m_1 общий знаменатель коэффициентов $g(x)$ и через m_2 общий знаменатель коэффициентов $h(x)$. Тогда

$$g(x)=\frac{1}{m_1}g_1(x) \quad \text{и} \quad h(x)=\frac{1}{m_2}h_1(x),$$

где $g_1(x)$ и $h_1(x)$ — многочлены уже с целыми коэффициентами. Далее, обозначим через d_1 наибольший общий делитель коэффициентов $g_1(x)$ и через d_2 наибольший общий делитель коэффициентов $h_1(x)$. Мы можем в таком случае написать, что

$$g(x)=\frac{1}{m_1}g_1(x)=\frac{d_1}{m_1}\varphi(x),$$

$$h(x)=\frac{1}{m_2}h_1(x)=\frac{d_2}{m_2}\psi(x),$$

где $\varphi(x)$ и $\psi(x)$ — примитивные многочлены. Отсюда

$$f(x)=\frac{d_1d_2}{m_1m_2}\varphi(x)\psi(x),$$

или, полагая $\frac{d_1d_2}{m_1m_2}=\frac{r}{s}$, где $\frac{r}{s}$ — несократимая дробь:

$$f(x)=\frac{r}{s}\varphi(x)\psi(x). \tag{7}$$

Если теперь

$$\varphi(x)\psi(x)=c_0+c_1x+\ldots+c_nx^n,$$

то из равенства (7) следует, что

$$a_0=\frac{rc_0}{s}, \quad a_1=\frac{rc_1}{s}, \ldots, a_n=\frac{rc_n}{s}.$$

Так как a_0 — целое число, то rc_0 должно делиться на s. Но в силу несократимости дроби $\frac{r}{s}$ числа r и s взаимно просты. Следовательно, c_0 должно делиться на s. Точно так же находим, что $c_1, c_2, \ldots, c_n$ делятся на s. Мы видим отсюда, что s есть общий делитель коэффициентов произведения $\varphi(x)\psi(x)$. По лемме 1 многочлен $\varphi(x)\psi(x)$ должен быть примитивным. Тем самым s должно равняться единице и потому

$$f(x)=r\varphi(x)\psi(x),$$

т. е. мы получили разложение многочлена $f(x)$ в произведение многочленов с целыми коэффициентами выше нулевой степени.

Теперь мы можем приступить к изложению основной идеи метода разложения произвольного многочлена на неприводимые множители над полем рациональных чисел. Для большей наглядности обратимся к конкретному примеру.

Пример 8. Разложить многочлен

$$f(x)=x^5+x^4+3x^3+4x^2+4x+2$$

на множители в поле рациональных чисел.

Легко убедиться, что данный многочлен $f(x)$ рациональных корней не имеет. Таким образом, если многочлен $f(x)$ приводим в поле рациональных чисел, то он должен разлагаться на два множителя второй и третьей степени. По лемме 2 коэффициенты этих множителей должны быть целыми числами, причём старшие коэффициенты должны равняться единице, так как старший коэффициент $f(x)$ равен единице. Следовательно, если многочлен $f(x)$ приводим, то его множитель второй степени должен иметь вид

$$g(x)=x^2+px+q,$$

где p, q — какие-то целые числа. Отсюда получается, что при любом целом m число $f(m)$ должно делиться на $g(m)$. Этим обстоятельством мы и воспользуемся для нахождения многочлена $g(x)$.

Так как $f(0)=2$ и $f(-1)=-1$, то для $g(0)$ и $g(-1)$ возможны только следующие комбинации значений:

1) $g(0)=1$ и $g(-1)=1$, 5) $g(0)=2$ и $g(-1)=1$,
2) $g(0)=1$ и $g(-1)=-1$, 6) $g(0)=2$ и $g(-1)=-1$,
3) $g(0)=-1$ и $g(-1)=1$, 7) $g(0)=-2$ и $g(-1)=-1$,
4) $g(0)=-1$ и $g(-1)=-1$, 8) $g(0)=-2$ и $g(-1)=1$.

Рассмотрим первую комбинацию $g(0)=1$ и $g(-1)=1$. Имеем:

$$g(0)=q=1,\quad g(-1)=1-p+q=1,$$

откуда

$$p=q=1 \quad \text{и} \quad g(x)=x^2+x+1.$$

Делим $f(x)$ на x^2+x+1 и убеждаемся, что $f(x)$ на этот квадратный трёхчлен делится:

$$f(x)=(x^2+x+1)(x^3+2x+2).$$

Таким образом, наша цель достигнута — разложение $f(x)$ в поле рациональных чисел на неприводимые множители получено, и дальнейшие комбинации 2) — 8) незачем рассматривать.

Во многих случаях большую помощь оказывают критерии неприводимости, позволяющие сразу обнаружить неприводимость многих многочленов в поле рациональных чисел. Укажем один из наиболее распространённых критериев.

Признак неприводимости Эйзенштейна. *Пусть $f(x)$ — многочлен с целыми коэффициентами. Если все его коэффициенты, кроме старшего, делятся на некоторое простое число p, а свободный член, делясь на p, не делится на p^2, то многочлен неприводим в поле рациональных чисел.*

Доказательство. Допустим противное — предположим, что многочлен

$$f(x)=a_0+a_1x+\ldots+a_nx^n$$

приводим в поле рациональных чисел. Тогда по лемме 2 многочлен $f(x)$ разложится в произведение двух многочленов $g(x)$ и $h(x)$ низшей степени с целыми коэффициентами:

$$f(x)=g(x)\,h(x). \tag{8}$$

Пусть

$$g(x)=b_0+b_1x+\ldots+b_kx^k \qquad (b_k\neq 0,\ 0<k<n),$$
$$h(x)=c_0+c_1x+\ldots+c_lx^l \qquad (c_l\neq 0,\ 0<l<n).$$

Тогда из равенства (8) следует, что

$$\begin{aligned}
a_0&=b_0c_0,\\
a_1&=c_0b_1+c_1b_0,\\
a_2&=c_0b_2+c_1b_1+c_2b_0,\\
&\ldots\ldots\ldots\ldots\\
a_k&=c_0b_k+c_1b_{k-1}+\ldots+c_kb_0,\\
&\ldots\ldots\ldots\ldots\ldots\\
a_n&=b_kc_l.
\end{aligned}$$

По условию свободный член $a_0=b_0c_0$ многочлена $f(x)$ делится на простое число p. Отсюда на p должно делиться b_0 или c_0. Но b_0 и c_0 не могут одновременно делиться на p, так как a_0 не делится на p^2. Пусть для определённости b_0 делится на p, но c_0 не делится на p. Возьмём равенство

$$a_1=c_0b_1+c_1b_0.$$

Его левая часть a_1 по условию делится на p, а в правой части член c_1b_0 делится на p, так как в этот член входит b_0, делящееся на p. Отсюда следует, что другой член в правой части c_0b_1 делится на p. Но c_0 не делится на простое число p. Поэтому b_1 должно делиться на p. Переходя к следующему равенству

$$a_2=c_0b_2+c_1b_1+c_2b_0,$$

подобным же образом убеждаемся, что b_2 делится на p и т. д. Наконец, из равенства

$$a_k=c_0b_k+c_1b_{k-1}+\ldots+c_kb_0$$

вытекает, что b_k делится на p.

Теперь обратимся к равенству $a_n=b_kc_l$. Так как b_k делится на p, то a_n должно делиться на p. Получилось противоречие с условиями признака, согласно которым старший коэффициент a_n

многочлена $f(x)$ не делится на p. Этим справедливость признака и доказана.

Пример 9. Многочлен

$$f(x)=x^5-12x^3+36x^2-12x-12$$

неприводим в поле рациональных чисел. В самом деле, для простого числа $p=3$ все условия признака Эйзенштейна выполняются — старший коэффициент рассматриваемого многочлена не делится на 3, а остальные коэффициенты делятся на 3, причём свободный член — 12 не делится на $3^2=9$.

С помощью признака Эйзенштейна можно построить многочлен произвольной степени, неприводимый в поле рациональных чисел, например

$$f(x)=x^n+2x^{n-1}+2x^{n-2}+\ldots+2.$$

В некоторых случаях, когда признак Эйзенштейна непосредственно применить не удаётся, можно, полагая $x=\alpha y+\beta$, где α и β — надлежащим образом подобранные рациональные числа, получить многочлен от переменного y

$$f_1(y)=f(\alpha y+\beta)=f(x),$$

удовлетворяющий условиям признака Эйзенштейна. В этом случае из неприводимости многочлена $f_1(x)$ сразу будет вытекать неприводимость $f(x)$. Действительно, если бы $f(x)$ был бы приводим, в то время как $f_1(x)$ неприводим, то

$$f(x)=g(x)\,h(x)$$

или

$$f(\alpha y+\beta)=f_1(y)=g(\alpha y+\beta)\;h(\alpha y+\beta)=g_1(y)\,h_1(y),$$

т. е. $f_1(x)$ было бы, вопреки условию, приводимым.

Пример 10. С помощью признака Эйзенштейна показать, что многочлен

$$f(x)=x^4-x^3+2x+1$$

неприводим в поле рациональных чисел.

К этому многочлену признак Эйзенштейна непосредственно применить нельзя, так как нельзя подобрать простого числа p, для которого будут выполняться все условия признака. Поэтому положим $x=y+1$. Тогда получим:

$$f(y+1)=f_1(y)=(y+1)^4-(y+1)^3+2(y+1)+1=$$
$$=y^4+3y^3+3y^2+3y+3.$$

При $p=3$ для многочлена $f_1(y)$ все условия признака Эйзенштейна выполняются. Следовательно, $f_1(y)$ и потому $f(x)$ неприводимы в поле рациональных чисел.

§ 6. Основная теорема алгебры

В произвольном числовом поле P многочлен n-й степени $f(x)$, как мы знаем, имеет не более чем n корней (считая каждый корень столько раз, какова его кратность) [1]). Однако многочлен $f(x)$ в поле P может и совсем не иметь корней. Возникает естественный вопрос, в каком числовом поле любой многочлен n-й степени имеет в точности n корней. Мы покажем, что таким полем является поле комплексных чисел, самое обширное числовое поле. Это замечательное свойство поля комплексных чисел вытекает из теоремы, носящей название основной теоремы алгебры, и заключающейся в том, что всякий многочлен $f(x)$ степени $n \geqslant 1$ с комплексными коэффициентами имеет в поле комплексных чисел по меньшей мере один корень.

Займёмся предварительно изучением основных свойств понятия непрерывности функции в области комплексных чисел, так как это понятие будет играть существенную роль в нашем доказательстве основной теоремы алгебры.

Обычное определение непрерывной функции, излагаемое в курсе математического анализа, можно без особых затруднений перенести и на тот случай, когда переменное x и функция $\varphi(x)$ принимают комплексные значения. А именно, назовём однозначную комплексную функцию $\varphi(x)$ комплексного переменного x *непрерывной в точке* x_0, если для всякого наперёд заданного положительного числа ε можно указать такое положительное число δ, что для всякого значения x, удовлетворяющего неравенству

$$|x - x_0| < \delta, \tag{1}$$

будет иметь место неравенство

$$|\varphi(x) - \varphi(x_0)| < \varepsilon \tag{2}$$

для соответствующих значений функции.

Обращаем внимание читателя на то, что здесь две вертикальные чёрточки означают уже не абсолютную величину, а модуль.

Для большей наглядности дадим геометрическое истолкование этому определению непрерывности. Возьмём для изображения значений переменного x и функции $\varphi(x)$ две плоскости P и Q. На плоскости P выберем прямоугольную систему координат $O\xi\eta$, а на плоскости Q — прямоугольную систему координат Ouv. Плоскость P мы будем называть плоскостью переменного x, а Q — плоскостью

---

[1]) Числовым полем принято называть всякую часть поля комплексных чисел, также образующую поле относительно арифметических действий сложения и умножения. Эта часть может совпадать и со всем полем комплексных чисел.

функции. Так как x и $\varphi(x)$ принимают комплексные значения, то можно написать, что

$$x=\xi+i\eta,\quad \varphi(x)=u+iv,$$

где ξ, η, u, v действительны, а $i=\sqrt{-1}$ — мнимая единица.

Каждое значение $x=\xi+i\eta$ можно на плоскости P изобразить точкой X с координатами ξ, η или вектором $\overrightarrow{OX}$.

Соответствующее значение функции $w=\varphi(x)=u+iv$ будет тогда изображаться точкой W с координатами u, v или вектором $\overrightarrow{OW}$ на плоскости Q.

Пусть теперь на плоскости P переменного x значение x_0 представляется вектором $\overrightarrow{OX_0}$. Тогда модуль разности $x-x_0$ будет означать расстояние между точками X и X_0. Таким образом, неравенство (1) означает, что точка X должна лежать внутри круга C_1 радиуса δ и с центром в точке X_0. Точно такое же истолкование можно дать неравенству (2): если на плоскости функции $\varphi(x)$ значение $\varphi(x_0)$ изображается точкой W_0, то неравенство (2) означает, что точка W лежит внутри круга C_2 радиуса ε и с центром в точке W_0.

Теперь определению непрерывности функции в точке можно дать следующее наглядное истолкование: функция $\varphi(x)$ непрерывна в точке x_0, если для круга C_2 произвольного радиуса ε с центром в точке W_0 на плоскости функции можно указать на плоскости переменного x круг C_1 такого радиуса δ с центром в точке X_0, что всякой внутренней точке круга C_1 будет соответствовать внутренняя точка круга C_2.

Если функция $\varphi(x)$ непрерывна в любой точке плоскости переменного x, то функция $\varphi(x)$ называется *всюду непрерывной* или, короче, *непрерывной*.

Посмотрим теперь, что можно сказать относительно многочлена с комплексными коэффициентами, если его рассматривать как комплексную функцию комплексного переменного [1]).

Легко убедиться в справедливости следующего предложения:

Многочлен $f(x)$ с комплексными коэффициентами есть непрерывная функция комплексного переменного x.

Доказательство. Пусть

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n$$

и x_0 — произвольное значение x или, выражаясь геометрически,

---

[1]) Так как поле комплексных чисел есть частный случай бесконечной области целостности R, то мы здесь имеем полное право рассматривать многочлены над полем комплексных чисел как функции комплексного переменного x.

произвольная точка. Так как

$$f(x_0) = a_0 x_0^n + a_1 x_0^{n-1} + \ldots + a_n,$$

то

$$f(x) - f(x_0) = a_0 (x^n - x_0^n) + a_1 (x^{n-1} - x_0^{n-1}) + \ldots + a_{n-1} (x - x_0).$$

Отсюда, принимая во внимание, что

$$x^k - x_0^k = (x - x_0)(x^{k-1} + x^{k-2} x_0 + \ldots + x_0^{k-1}),$$

получаем:

$$f(x) - f(x_0) = (x - x_0)[a_0 (x^{n-1} + x^{n-2} x_0 + \ldots + x_0^{n-1}) + \ldots + a_{n-1}].$$

Известно, что модуль суммы меньше или равен сумме модулей, а модуль произведения равен произведению модулей. Следовательно,

$$|f(x) - f(x_0)| \leqslant |x - x_0| [|a_0| (|x|^{n-1} + |x|^{n-2} |x_0| + \ldots + |x_0|^{n-1}) + \ldots + |a_{n-1}|]. \quad (3)$$

Возьмём точку x столь близкой к точке x_0, чтобы x лежало внутри круга радиуса 1 с центром в точке x_0, т. е. чтобы $|x - x_0| < 1$.

Тогда $|x| < |x_0| + 1$. Так как, очевидно, и $|x_0| < |x_0| + 1$, то, полагая для сокращения письма $|x_0| + 1 = M$, мы можем неравенство (3) усилить, заменяя в нём $|x|$ и $|x_0|$ большей величиной M. Получим:

$$|f(x) - f(x_0)| < |x - x_0| [|a_0| \underbrace{(M^{n-1} + \ldots + M^{n-1})}_{n \text{ раз}} + |a_1| \underbrace{(M^{n-2} + \ldots + M^{n-2})}_{n-1 \text{ раз}} + \ldots + |a_{n-1}|]$$

или

$$|f(x) - f(x_0)| < |x - x_0| [n |a_0| M^{n-1} + (n-1) |a_1| M^{n-2} + \ldots + |a_{n-1}|]. \quad (4)$$

Обозначим

$$n |a_0| M^{n-1} + (n-1) |a_1| M^{n-2} + \ldots + |a_{n-1}|$$

через N. Тогда неравенство (4) запишется в виде

$$|f(x) - f(x_0)| < N |x - x_0|. \quad (5)$$

А теперь возьмём для заданного $\varepsilon > 0$ число $\delta > 0$ так, чтобы $\delta < \frac{\varepsilon}{N}$ и $\delta < 1$. Полагая $|x - x_0| < \delta$, мы получим следующее усиление неравенства (5):

$$|f(x) - f(x_0)| < N \frac{\varepsilon}{N}$$

или окончательно

$$|f(x) - f(x_0)| < \varepsilon.$$

Итак, многочлен $f(x)$ оказался функцией, непрерывной в точке x_0. Но x_0 — произвольно. Следовательно, $f(x)$ есть всюду непрерывная функция, и наше утверждение доказано.

Из непрерывности многочлена вытекает, что и *его модуль* $|f(x)|$ *есть также непрерывная функция комплексного переменного* x.

В самом деле, согласно одному из свойств модуля мы можем написать, что

$$||f(x)|-|f(x_0)|| \leqslant |f(x)-f(x_0)|. \tag{6}$$

Так как по доказанному многочлен $f(x)$ есть непрерывная функция комплексного переменного x, то для любого $\varepsilon > 0$ можно указать такое $\delta > 0$, что при $|x-x_0| < \delta$ будет выполняться неравенство

$$|f(x)-f(x_0)| < \varepsilon.$$

Отсюда благодаря неравенству (6) и подавно будет:

$$||f(x)|-|f(x_0)|| < \varepsilon$$

при $|x-x_0| < \delta$.

Из теоремы о непрерывности многочлена вытекает ещё одно следствие:

Пусть

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x$$

— многочлен степени $n \geqslant 0$ *и со свободным членом* a_n, *равным нулю. Тогда для любого* $\varepsilon > 0$ *можно указать такое* $\delta > 0$, *что для всех* x, *удовлетворяющих неравенству* $|x| < \delta$, *будет иметь место неравенство* $|f(x)| < \varepsilon$.

В самом деле, возьмём в качестве x_0 значение 0. В силу непрерывности многочлена $f(x)$ можно для любого $\varepsilon > 0$ указать такое $\delta > 0$, что для всех x, удовлетворяющих условию $|x-0| < \delta$, будет выполняться неравенство

$$|f(x)-f(0)| < \varepsilon.$$

Но $f(0) = 0$, так как $f(x)$ есть многочлен со свободным членом, равным нулю. Таким образом, получается, что для любого $\varepsilon > 0$ можно указать такое $\delta > 0$, для которого $|f(x)| < \varepsilon$ при $|x| < \delta$.

Для доказательства основной теоремы алгебры понадобится, кроме того, несколько лемм.

Лемма 1. *Модуль всякого многочлена*

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

степени $n \geqslant 1$ *при достаточно больших по модулю значениях* x *может быть сделан больше любого наперёд заданного действительного положительного числа* M.

Доказательство. Преобразуем многочлен $f(x)$ следующим образом:

$$f(x)=a_0x^n\left[1+\left(\frac{a_1}{a_0}\cdot\frac{1}{x}+\frac{a_2}{a_0}\cdot\frac{1}{x^2}+\ldots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}\right)\right].$$

Так как модуль суммы больше или равен разности модулей, то

$$|f(x)|\geqslant|a_0|\cdot|x|^n\left[1-\left|\frac{a_1}{a_0}\cdot\frac{1}{x}+\ldots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}\right|\right]. \tag{7}$$

Выражение

$$\frac{a_1}{a_0}\cdot\frac{1}{x}+\frac{a_2}{a_0}\cdot\frac{1}{x^n}+\ldots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}$$

можно рассматривать как многочлен от $\frac{1}{x}$ со свободным членом, равным нулю. Поэтому можно для $\varepsilon=\frac{1}{2}$ указать такое $\delta>0$, что при $\left|\frac{1}{x}\right|<\delta$ будет иметь место неравенство

$$\left|\frac{a_1}{a_0}\cdot\frac{1}{x}+\ldots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}\right|<\frac{1}{2}. \tag{8}$$

Иными словами, неравенство (8) будет выполняться при $|x|>\frac{1}{\delta}=N$. Отсюда, полагая $|x|>N$, можно неравенство (7) усилить, а именно,

$$|f(x)|>|a_0|\cdot|x|^n\left[1-\frac{1}{2}\right]$$

или

$$|f(x)|>\frac{1}{2}|a_0|\cdot|x|^n.$$

Возьмём теперь x столь большим по модулю, чтобы одновременно $|x|>N$ и $|x|>N_1$, где

$$N_1=\sqrt[n]{\frac{2M}{|a_0|}}.$$

Тогда

$$|f(x)|>\frac{1}{2}|a_0|\cdot|x|^n>\frac{1}{2}|a_0|\left(\sqrt[n]{\frac{2M}{|a_0|}}\right)^n$$

или $|f(x)|>M$, что и требовалось показать.

Мы подошли к лемме, играющей в доказательстве основной теоремы алгебры весьма важную роль.

Лемма 2 (лемма Даламбера). *Если многочлен*

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n$$

степени $n\geqslant 1$ *не обращается в нуль при* $x=x_0$, *то можно подобрать такое комплексное число* h, *чтобы*

$$|f(x_0+h)|<|f(x_0)|.$$

Доказательство. Положим $x = x_0 + h$. Тогда

$$f(x_0 + h) = a_0 (x_0 + h)^n + a_1 (x_0 + h)^{n-1} + \ldots + a_n.$$

Раскрывая степени $(x_0 + h)^k$ по биному Ньютона и располагая члены по возрастающим степеням h, получаем:

$$f(x_0 + h) = b_0 + b_1 h + b_2 h^2 + \ldots + b_{n-1} h^{n-1} + a_0 h^n, \qquad (9)$$

где $b_0, b_1, \ldots, b_{n-1}$ — соответствующие комплексные числа. Полагая в равенстве (9) $h = 0$, находим, что $b_0 = f(x_0)$. Таким образом,

$$f(x_0 + h) = f(x_0) + b_1 h + b_2 h^2 + \ldots + b_{n-1} h^{n-1} + a_0 h^n. \qquad (10)$$

Может случиться, однако, что часть или даже все коэффициенты $b_1, b_2, \ldots, b_{n-1}$ будут равны нулю. Рассмотрим сначала тот случай, когда не все b_i равны нулю. В этом случае можно указать такое число k $(1 \leqslant k < n)$, что $b_k \neq 0$, но $b_1 = b_2 = \ldots = b_{k-1} = 0$ (в частности, если $b_1 \neq 0$, то $k = 1$). В этом предположении равенство (10) перепишется следующим образом:

$$f(x_0 + h) = f(x_0) + b_k h^k + b_{k+1} h^{k+1} + \ldots + a_0 h^n.$$

Разделим обе части последнего равенства на $f(x_0)$. Делить на $f(x_0)$ мы имеем право, так как по условию $f(x_0) \neq 0$. Получим:

$$\frac{f(x_0 + h)}{f(x_0)} = 1 + c_k h^k + c_{k+1} h^{k+1} + \ldots + c_n h^n, \qquad (11)$$

где

$$c_k = \frac{b_k}{f(x_0)}, \quad \ldots, \quad c_{n-1} = \frac{b_{n-1}}{f(x_0)}, \quad c_n = \frac{a_0}{f(x_0)},$$

причём $c_k \neq 0$, так как, по нашему предположению, $b_k \neq 0$.

Принимая во внимание, что $c_k \neq 0$, мы можем правую часть равенства (11) преобразовать так:

$$\frac{f(x_0 + h)}{f(x_0)} = (1 + c_k h^k) + c_k h^k \left(\frac{c_{k+1}}{c_k} h + \ldots + \frac{c_n}{c_k} h^{n-k} \right).$$

Воспользуемся теперь тем, что модуль суммы меньше или равен сумме модулей и модуль произведения равен произведению модулей:

$$\left| \frac{f(x_0 + h)}{f(x_0)} \right| \leqslant |1 + c_k h^k| + |c_k h^k| \cdot \left| \frac{c_{k+1}}{c_k} h + \ldots + \frac{c_n}{c_k} h^{n-k} \right|. \qquad (12)$$

Выражение

$$\frac{c_{k+1}}{c_k} h + \ldots + \frac{c_n}{c_k} h^{n-k}$$

можно рассматривать как многочлен от h со свободным членом, равным нулю. Для такого многочлена можно при $\varepsilon = \frac{1}{2}$ указать такое $\delta > 0$, что при $|h| < \delta$ будет:

$$\left| \frac{c_{k+1}}{c_k} h + \ldots + \frac{c_n}{c_k} h^{n-k} \right| < \frac{1}{2}.$$

Пользуясь этим обстоятельством, усиливаем неравенство (12), предполагая $|h|<\delta$:

$$\left|\frac{f(x_0+h)}{f(x_0)}\right|<|1+c_kh^k|+\frac{1}{2}|c_kh^k|. \tag{13}$$

До сих пор мы подбирали h достаточно малым по модулю. Теперь подберем аргумент h таким, чтобы c_kh^k было действительным отрицательным числом. Для этого надо потребовать, чтобы $\arg(c_kh^k)=\pi$, так как аргумент отрицательного действительного числа равен π. Отсюда получаем, что $\arg c_k+k\arg h=\pi$ и

$$\arg h=\frac{\pi-\arg c_k}{k}. \tag{14}$$

С этого момента мы будем всё время предполагать аргумент выбранным таким, как это указано в равенстве (14).

Теперь при указанном выборе аргумента h имеем, что $c_kh^k=$ $=-|c_kh^k|$, и неравенство (13) можно переписать следующим образом:

$$\left|\frac{f(x_0+h)}{f(x_0)}\right|<|1-|c_kh^k||+\left|\frac{1}{2}\left|c_kh^k\right|\right|.$$

Может случиться, что $|c_kh^k|$ будет больше единицы. В таком случае возьмём h по модулю столь малым, чтобы $|c_kh^k|<1$. Тогда разность $1-|c_kh^k|$ будет положительна и потому

$$|1-|c_kh^k||=1-|c_kh^r|.$$

Отсюда

$$\left|\frac{f(x_0+h)}{f(x_0)}\right|<1-|c_kh^k|+\frac{1}{2}\left|c_kh^k\right|=1-\frac{1}{2}|c_kh^k|.$$

Но при $|c_kh^k|<1$ неравенство $1-\frac{1}{2}|c_kh^k|<1$ будет, очевидно, выполняться. Следовательно, и подавно

$$\left|\frac{f(x_0+h)}{f(x_0)}\right|<1$$

при указанном выборе h по модулю и аргументу. Так как

$$\left|\frac{f(x_0+h)}{f(x_0)}\right|=\frac{|f(x_0+h)|}{|f(x_0)|},$$

то отсюда следует, что $\frac{|f(x_0+h)|}{|f(x_0)|}<1$, откуда

$$|f(x_0+h)|<|f(x_0)|.$$

Остаётся разобрать случай, когда все коэффициенты b_i выражения (10) равны нулю. Но в этом случае дело обстоит ещё проще. А именно, если $b_1=b_2=\ldots=b_{n-1}=0$, то

$$f(x_0+h)=f(x_0)+a_0h^n,$$

откуда

$$\frac{f(x_0+h)}{f(x_0)} = 1 + c_n h^n,$$

где $c_n = \frac{a_0}{f(x_0)} \neq 0$. Таким образом,

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| = |1 + c_n h^n|.$$

Полагая затем $\arg h = \frac{\pi - \arg c_n}{n}$ и $|c_n h^n| < 1$, получаем, что

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| = |1 - |c_n h^n|| = 1 - |c_n h^n| < 1,$$

откуда

$$|f(x_0+h)| < |f(x_0)|.$$

Итак, лемма доказана полностью.

Кроме леммы Даламбера, нам также понадобятся некоторые свойства последовательностей комплексных чисел, связанных с понятием предела. При этом мы будем предполагать, что читатель уже имеет некоторое знакомство с понятием предела и сходимости числовых последовательностей в области действительных чисел.

Пусть

$$x_1, x_2, \ldots, x_k, \ldots \tag{15}$$

— некоторая последовательность комплексных чисел, $x_k = a_k + ib_k$, где a_k, b_k — действительные числа. Комплексное число α называется *пределом числовой последовательности* (15), если для любого $\varepsilon > 0$ можно указать такое число $N > 0$, что $|x_k - \alpha| < \varepsilon$ для всех $k > N$. При этом последовательность (15) называется *сходящейся* к числу α. То обстоятельство, что α есть предел последовательности (15), мы будем коротко записывать так: $\lim\limits_{k\to\infty} x_k = \alpha$. Последовательность, не имеющая предела, называется *расходящейся*. Для сокращения письма мы будем последовательность (15) обозначать через $\{x_k\}$.

Покажем, что *последовательность* $\{x_k = a_k + ib_k\}$ *сходится тогда и только тогда к числу* $\alpha = \xi + i\eta$, *когда последовательности действительных чисел* $\{a_k\}$ *и* $\{b_k\}$ *сходятся соответственно к числам* ξ *и* η.

Доказательство. Если $\lim\limits_{k\to\infty} x_k = \alpha$, то согласно определению предела числовой последовательности можно для всякого $\varepsilon > 0$ указать такое положительное число N, что для всех $k > N$

$$|x_k - \alpha| < \varepsilon. \tag{16}$$

Так как $x_k - \alpha = (a_k - \xi) + i(b_k - \eta)$, то неравенство (16) можно переписать следующим образом:

$$\sqrt{(a_k - \xi)^2 + (b_k - \eta)^2} < \varepsilon,$$

откуда и подавно

$$|a_k-\xi|<\varepsilon, \quad |b_k-\eta|<\varepsilon.$$

Мы видим, что последовательность $\{a_k\}$ сходится к ξ, а последовательность $\{b_k\}$ сходится к η: $\lim\limits_{k\to\infty} a_k=\xi$, $\lim\limits_{k\to\infty} b_k=\eta$.

Обратно, пусть $\lim\limits_{k\to\infty} a_k=\xi$ и $\lim\limits_{k\to\infty} b_k=\eta$. Тогда для всякого $\varepsilon_1=\frac{\varepsilon}{\sqrt{2}}>0$ можно указать такие положительные числа N_1 и N_2, что при $k>N_1$

$$|a_k-\xi|<\frac{\varepsilon}{\sqrt{2}} \tag{17}$$

и

$$|b_k-\eta|<\frac{\varepsilon}{\sqrt{2}}. \tag{18}$$

Таким образом, если N — наибольшее из чисел N_1 и N_2, то неравенства (17) и (18) будут одновременно выполняться для всех $k>N$. Отсюда при $k>N$ получаем, что

$$|x_k-\alpha|=\sqrt{(a_k-\xi)^2+(b_k-\eta)^2}<\sqrt{\frac{\varepsilon^2}{2}+\frac{\varepsilon^2}{2}}=\varepsilon,$$

т. е. последовательность $\{x_k\}$ сходится к числу α.

На основании только что доказанного можно вывести следующее свойство непрерывных функций комплексного переменного: *если комплексная функция $\varphi(x)$ комплексного переменного непрерывна в точке x_0, то для любой последовательности $\{x_k\}$ комплексных чисел, сходящейся к x_0, будет иметь место равенство*

$$\lim_{k\to\infty}\varphi(x_k)=\varphi(x_0).$$

В самом деле, согласно определению непрерывности в точке для всякого числа $\varepsilon>0$ можно указать такое $\delta>0$, что для всех x, удовлетворяющих условию $|x-x_0|<\delta$, будет выполняться неравенство $|\varphi(x)-\varphi(x_0)|<\varepsilon$.

Пусть теперь $\{x_k\}$ — последовательность комплексных чисел, сходящаяся к x_0. Тогда для $\delta>0$ можно указать такое $N>0$, что $|x_k-x_0|<\delta$ при $k>N$. Для этих чисел x_k, очевидно, будет иметь место неравенство

$$|\varphi(x_k)-\varphi(x_0)|<\varepsilon.$$

Мы видим отсюда, что последовательность $\{\varphi(x_k)\}$ сходится к $\varphi(x_0)$, т. е. $\lim\limits_{k\to\infty}\varphi(x_k)=\varphi(x_0)$.

Назовём последовательность комплексных чисел $\{x_k\}$ (сходящуюся или расходящуюся — безразлично) *ограниченной*, если можно указать такое достаточно большое положительное число M, что $|x_k|<M$ $(k=1, 2, 3, \ldots)$.

Назовём, далее, *подпоследовательностью* последовательности комплексных чисел $\{x_k\}$ последовательность $\{x_{\nu_k}\}$, где $\nu_1, \nu_2, \ldots, \nu_k, \ldots$ есть некоторая монотонно возрастающая последовательность положительных целых чисел: $\nu_1 < \nu_2 < \ldots < \nu_k < \ldots$ Нетрудно убедиться, что *если последовательность* $\{x_k\}$ *сходится к* x_0, *то и всякая её подпоследовательность* $\{x_{\nu_k}\}$ *сходится к* x_0.

В самом деле, так как $\{x_k\}$ сходится к x_0, то для любого $\varepsilon > 0$ можно указать такое $N > 0$, что при $k > N$ будет иметь место неравенство $|x_k - x_0| < \varepsilon$. Отсюда для всех $\nu_k > N$ будет также $|x_{\nu_k} - x_0| < \varepsilon$, т. е. подпоследовательность $\{x_{\nu_k}\}$ сходится к тому же числу x_0.

Отметим следующее важное свойство ограниченных последовательностей:

Всякая ограниченная последовательность комплексных чисел $\{x_k\}$ *обладает сходящейся подпоследовательностью.*

Доказательство. Мы будем считать эту теорему известной для последовательностей действительных чисел. В соответствии с этим мы проведём доказательство так.

Прежде всего легко усмотреть, что из ограниченности последовательности $\{x_k = a_k + ib_k\}$ вытекает ограниченность последовательностей действительных чисел $\{a_k\}$ и $\{b_k\}$. В самом деле, так как $|a_k| \leqslant |x_k|$ и $|b_k| \leqslant |x_k|$, то из неравенства $|x_k| < M$ следует, что и подавно $|a_k| < M$, $|b_k| < M$.

Поскольку последовательность действительных чисел $\{a_k\}$ ограничена, она должна обладать сходящейся подпоследовательностью. Пусть это будет $\{a_{\nu_k}\}$. Тогда $\{x_{\nu_k} = a_{\nu_k} + ib_{\nu_k}\}$ будет некоторой подпоследовательностью последовательности $\{x_k\}$. Рассмотрим последовательность $\{b_{\nu_k}\}$ мнимых частей чисел x_{ν_k}. Эта последовательность есть подпоследовательность ограниченной последовательности $\{b_k\}$ и потому в свою очередь ограничена. Следовательно, $\{b_{\nu_k}\}$ должна обладать сходящейся подпоследовательностью. Пусть это будет $\{b_{\mu_k}\}$. Рассмотрим последовательность $\{x_{\mu_k} = a_{\mu_k} + ib_{\mu_k}\}$. Она сходится, так как $\{b_{\mu_k}\}$ сходится, и $\{a_{\mu_k}\}$ сходится (как подпоследовательность сходящейся последовательности $\{a_{\nu_k}\}$). Мы видим отсюда, что $\{x_k\}$ обладает сходящейся подпоследовательностью $\{x_{\mu_k}\}$.

Вернёмся теперь к многочлену $f(x)$ и рассмотрим множество A всевозможных значений модуля многочлена $|f(x)|$. Так как модуль комплексного числа не может быть отрицательным, то $|f(x)| \geqslant 0$. Таким образом, это множество A оказывается ограниченным снизу. Известно, что всякое (не пустое) множество действительных чисел, ограниченное снизу, должно обладать нижней гранью. Следовательно, и для множества A должна существовать нижняя грань, т. е. должно существовать такое действительное число l, что $|f(x)| \geqslant l$ для

всех значений x, и для любого $\delta > 0$ можно указать такое значение $x = x'$, при котором $|f(x')| < l + \delta$. Иными словами, для любого $\delta > 0$ можно подобрать такое комплексное число x', что

$$0 \leqslant |f(x')| - l < \delta. \tag{19}$$

Покажем, что имеет место даже нечто большее, а именно справедлива следующая.

Теорема 15. *Если l — нижняя грань множества A всевозможных значений модуля многочлена $|f(x)|$, то существует по меньшей мере одно такое комплексное число x_0, что $l = |f(x_0)|$.*

Доказательство. Возьмём какую-нибудь последовательность $\{\delta_k\}$ положительных действительных чисел, сходящуюся к нулю. Согласно сказанному выше о неравенстве (19) можно для каждого δ_k подобрать такое x'_k, что

$$0 \leqslant |f(x'_k)| - l < \delta_k \qquad (k = 1, 2, 3, \ldots). \tag{20}$$

Последовательность $\{\delta_k\}$ сходится к нулю. Это значит, что для всякого $\varepsilon > 0$ можно указать такое $N > 0$, что для всех $k > N$ будет иметь место неравенство $\delta_k < \varepsilon$. Отсюда для $k > N$ можно неравенство (20) усилить, а именно при $k > N$ будет иметь место

$$0 \leqslant |f(x'_k)| - l < \varepsilon.$$

Последнее неравенство свидетельствует о том, что последовательность действительных положительных чисел $\{|f(x'_k)|\}$ сходится к l. Поскольку последовательность $\{|f(x'_k)|\}$ сходится, она должна быть ограниченной, т. е. должно существовать такое число $M > 0$, для которого

$$|f(x'_k)| < M \qquad (k = 1, 2, 3, \ldots). \tag{21}$$

Согласно лемме 1 можно для этого M указать такое положительное число N, что при $|x| > N$ будет выполняться неравенство

$$|f(x)| > M. \tag{22}$$

Но для чисел $x'_1, x'_2, \ldots, x'_k, \ldots$ имеет место неравенство (21), а не (22). Поэтому x'_k не могут быть по модулю больше N; иными словами, $|x'_k| \leqslant N$. Таким образом, последовательность $\{x'_k\}$ оказалась ограниченной. Но, как известно, ограниченная последовательность $\{x'_k\}$ должна обладать сходящейся подпоследовательностью. Пусть это будет $\{x'_{\nu_k}\}$, причём пусть $\lim\limits_{k \to \infty} x'_{\nu_k} = x_0$.

Так как $\{|f(x'_{\nu_k})|\}$ есть подпоследовательность последовательности $\{|f(x'_k)|\}$ сходящейся к l, то $\{|f(x'_{\nu_k})|\}$ также сходится к l. Следовательно, мы можем написать, что

$$\lim_{k \to \infty} |f(x'_{\nu_k})| = l. \tag{23}$$

Но в силу непрерывности модуля многочлена

$$\lim_{k\to\infty} |f(x'_{\nu_k})| = |f(x_0)|. \tag{24}$$

Таким образом, сравнивая равенства (23) и (24), получаем:

$$|f(x_0)| = l,$$

и теорема доказана.

Теперь мы можем сформулировать основную теорему алгебры и приступить к её доказательству.

Основная теорема алгебры. *Всякий многочлен $f(x)$ степени $n \geqslant 1$ с комплексными коэффициентами имеет в поле комплексных чисел по меньшей мере один корень.*

Доказательство. Как мы только что установили, существует по меньшей мере одно такое комплексное число x_0, для которого $|f(x_0)| = l$, где l — нижняя грань всех значений модуля многочлена $|f(x)|$. Покажем теперь, что $l = 0$. Допустив противное, предположим, что $l \neq 0$. Тогда $f(x_0) \neq 0$, и мы можем воспользоваться леммой Даламбера. Согласно этой лемме можно подобрать такое комплексное число $x' = x_0 + h$, что $|f(x')| < |f(x_0)|$ или $|f(x')| < l$. Но последнее неравенство противоречит тому, что l есть нижняя грань всех значений $|f(x)|$. Следовательно, наше допущение неверно и, таким образом, $l = |f(x_0)|$ должно равняться нулю. Отсюда $f(x_0) = 0$, т. е. x_0 оказалось корнем многочлена $f(x)$.

Из основной теоремы алгебры вытекает целый ряд следствий. Отметим наиболее существенные.

1°. *В поле комплексных чисел неприводимыми являются лишь многочлены первой степени.*

В самом деле, если $p(x)$ — многочлен, неприводимый в поле комплексных чисел, то он согласно основной теореме алгебры должен иметь по меньшей мере один комплексный корень x_0. Отсюда $p(x) = (x - x_0)\, q(x)$. Но в силу неприводимости $p(x)$ многочлен $q(x)$ не может иметь степень выше нулевой. Следовательно, $q(x) = c$, где c — некоторое комплексное число, отличное от нуля. Таким образом, $p(x) = c(x - x_0)$, т. е. $p(x)$ оказался многочленом первой степени, что и требовалось показать.

2°. *Всякий многочлен $f(x)$ степени $n \geqslant 1$ с комплексными коэффициентами разлагается в поле комплексных чисел целиком на линейные множители.*

Действительно, многочлен $f(x)$ должен следующим образом разлагаться на неприводимые множители в поле комплексных чисел

$$f(x) = a_0 (x - x_1)^{k_1} (x - x_2)^{k_2} \ldots (x - x_r)^{k_r}, \tag{25}$$

так как в этом поле неприводимыми являются лишь многочлены первой степени. При этом, очевидно, a_0 есть старший коэффициент многочлена $f(x)$.

3°. *Всякий многочлен $f(x)$ степени n с комплексными коэффициентами имеет n комплексных корней, считая при этом каждый корень столько раз, какова его кратность.* Действительно, это следствие с очевидностью вытекает из разложения (25).

Для получения дальнейших следствий воспользуемся следующим важным свойством многочлена с действительными коэффициентами: *если x_0 — корень многочлена $f(x)$ с действительными коэффициентами, то сопряжённое комплексное число $\bar{x}_0$ будет также корнем многочлена $f(x)$.* Иными словами, *мнимые корни многочлена $f(x)$ с действительными коэффициентами должны быть попарно сопряжены.*

Доказательство. Пусть $x_0 = \alpha + \beta i$ — мнимый корень многочлена

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n$$

с действительными коэффициентами. Обозначим через $\bar{x}_0$ комплексное число, сопряжённое с x_0, и разделим многочлен $f(x)$ на $g(x) = (x - x_0)(x - \bar{x}_0) = x^2 - 2\alpha x + (\alpha^2 + \beta^2)$. Так как степень делителя $g(x)$ равна двум, то остаток будет иметь вид $Px + Q$, причём его коэффициенты действительны, так как коэффициенты делимого $f(x)$ и делителя $g(x)$ действительны. Таким образом, можно написать, что

$$f(x) = (x - x_0)(x - \bar{x}_0)\, q(x) + (Px + Q).$$

Полагая в последнем равенстве $x = x_0$, получаем:

$$P(\alpha + \beta i) + Q = 0,$$

или

$$(\alpha P + Q) + i\beta P = 0,$$

откуда

$$\alpha P + Q = 0, \quad \beta P = 0.$$

Но $\beta \neq 0$, так как x_0 — мнимое число. Следовательно, $P = 0$ и потому $Q = 0$. Отсюда

$$f(x) = (x - x_0)(x - \bar{x}_0)\, q(x),$$

и теперь становится очевидным, что $\bar{x}_0$ есть также корень многочлена $f(x)$. Теперь укажем дальнейшие следствия из основной теоремы алгебры. Они будут относиться уже к многочленам с действительными коэффициентами.

4°. *В поле действительных чисел неприводимыми могут быть только многочлены не выше второй степени.*

В самом деле, пусть, вопреки утверждению, $f(x)$ есть многочлен с действительными коэффициентами выше второй степени и неприводимый в поле действительных чисел. Согласно основной теореме алгебры $f(x)$ должен иметь по меньшей мере один комп-

лексный корень x_0. Этот корень не может быть действительным: если бы x_0 было действительным, то многочлен $f(x)$ был бы приводимым в поле действительных чисел:

$$f(x)=(x-x_0)\,q(x),$$

где $q(x)$ — многочлен с действительными коэффициентами. Следовательно, x_0 есть мнимый корень. Отсюда в силу доказанного выше получаем при $x_0=\alpha+\beta i$ $(\beta\neq 0)$, что

$$f(x)=[x^2-2\alpha x+(\alpha^2+\beta^2)]\,q(x).$$

Так как коэффициенты квадратного трёхчлена $x^2-2\alpha x+(\alpha^2+\beta^2)$ действительны, то $q(x)$ должен быть многочленом также с действительными коэффициентами. Вместе с тем степень $q(x)$ не ниже единицы, так как $f(x)$ есть многочлен выше второй степени. Получилось противоречие с нашим предположением о неприводимости $f(x)$ в поле действительных чисел.

5°. *Всякий многочлен степени $n\geqslant 1$ с действительными коэффициентами разлагается в поле действительных чисел на множители не выше второй степени.*

6°. *Число действительных корней многочлена $f(x)$ степени $n\geqslant 1$ с действительными коэффициентами имеет ту же чётность, что и степень n многочлена.*

Справедливость этого свойства видна из следующих рассуждений.

Как мы уже знаем, многочлен $f(x)$ должен иметь n комплексных корней (см. 3°). Среди этих корней могут быть и действительные корни. Пусть число действительных корней многочлена $f(x)$ равно s $(0\leqslant s\leqslant n)$. Тогда $n-s$ будет числом мнимых корней многочлена. Из попарной сопряжённости мнимых корней следует, что их число $n-s$ должно быть чётным, откуда n и s имеют одинаковую чётность.

Из свойства 6°, в частности, вытекает, что *если степень многочлена $f(x)$ с действительными коэффициентами нечётна, то многочлен имеет нечётное число действительных корней и потому имеет по меньшей мере один действительный корень.*

Мысль о том, что многочлен n-й степени имеет n комплексных корней, возникла ещё в XVII в. и была высказана в 1629 г. французским математиком Жираром. В 1746 г. Даламбер сделал попытку доказать основную теорему алгебры. Первое удовлетворительное доказательство было дано в 1799 г. Гауссом. В настоящее время известно много различных доказательств этой теоремы [1]).

---

[1]) В учебниках Куроша [2] и Окунева [4] по курсу высшей алгебры приводится доказательство основной теоремы алгебры, связанное с понятием поля разложения. В книге Кузьмина-Фаддеева [1] можно найти доказательство, связанное с поведением аргумента многочлена $f(z)$ комплексного переменного z при обходе точки z замкнутого контура (см. стр. 41 — 51). Известны также и топологические доказательства основной теоремы алгебры. Одно из таких доказательств излагается на стр. 356 — 358 книги Р. Куранта и Г. Робинса «Что такое математика» (Гостехиздат, 1947).

В свое время эта теорема считалась краеугольным камнем всей алгебры. В настоящее время в связи с весьма интенсивным развитием таких разделов, как теория групп, теория колец и теория полей, эту теорему следует рассматривать только как основную теорему алгебры комплексных чисел.

§ 7. Проблема решения уравнений в радикалах. Двучленные уравнения

В предыдущем параграфе было установлено, что всякое алгебраическое уравнение n-й степени с комплексными коэффициентами

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0 \qquad (a_0 \neq 0) \tag{1}$$

имеет в точности n комплексных корней. Теперь возникает вопрос — как найти эти корни с помощью тех или иных операций, производимых над коэффициентами уравнения (1).

Основными алгебраическими операциями, совершаемыми над комплексными числами, являются четыре арифметических действия, а также действия возвышения в степень и извлечение корня. Поэтому является вполне естественной следующая задача, известная под названием проблемы решения уравнений в радикалах: *выразить корни уравнения* (1) *через его коэффициенты с помощью конечного числа действий сложения, вычитания, умножения, деления, возвышения в степень и извлечения корня.*

Операция извлечения корня n-й степени из комплексного числа a сводится к отысканию корней так называемого *двучленного уравнения*

$$x^n - a = 0 \qquad (a \neq 0). \tag{2}$$

В алгебре обычно под символом $\sqrt[n]{a}$ подразумевается один из корней двучленного уравнения (2), и этот символ часто называется *радикалом.*

Уже из курса элементарной алгебры известно, что уравнение второй степени решается в радикалах. В ближайших параграфах мы увидим, что решаются в радикалах и уравнения третьей и четвёртой степеней. Гораздо сложнее обстоит дело с уравнениями высших степеней. В § 16 мы покажем, что уравнения выше четвёртой степени, вообще говоря, нельзя решить в радикалах. Таким образом, алгебраические действия оказываются недостаточными для решения любого уравнения, имеющего степень выше четырёх.

Вернёмся к операции извлечения корня n-й степени из комплексного числа. Из теории комплексных чисел известно, что все n корней уравнения (2) можно найти по формуле

$$\sqrt[n]{a} = \left|\sqrt[n]{r}\right| \left(\cos\frac{\varphi + 2k\pi}{n} + i\sin\frac{\varphi + 2k\pi}{n}\right) \quad (k = 0, 1, \ldots, n-1), \tag{3}$$

где r — модуль a, φ — аргумент a и $\left|\sqrt[n]{r}\right|$ — арифметическое значение корня n-й степени из r.

Формула (3) допускает любопытное геометрическое истолкование. Возьмём для изображения комплексных чисел координатную плоскость P с прямоугольной системой координат $O\alpha\beta$. Тогда числа

$$x_k = \left|\sqrt[n]{r}\right|\left(\cos\frac{\varphi+2k\pi}{n} + i\sin\frac{\varphi+2k\pi}{n}\right)$$

будут изображаться на плоскости P векторами, имеющими начало в точке O, а концы — в вершинах правильного n-угольника, вписанного в окружность радиуса $R=\left|\sqrt[n]{r}\right|$ и с центром в точке O. Мы видим отсюда, что решение двучленного уравнения (2) тесно связано с задачей построения правильного n-угольника.

Пример. Найти правильный шестиугольник, соответствующий корням двучленного уравнения $x^6 - i = 0$, где $i = \sqrt{-1}$ — мнимая единица.

Так как модуль i равен 1, аргумент i равен $\frac{\pi}{2}$ и $n=6$, то по формуле (3) получаем:

$$\sqrt[6]{i} = \cos\frac{\frac{\pi}{2}+2k\pi}{6} + i\sin\frac{\frac{\pi}{2}+2k\pi}{6}.$$

Полагая k равным последовательно 0, 1, 2, 3, 4, 5, будем иметь все шесть корней уравнения

$$x_0 = \cos\frac{\pi}{12} + i\sin\frac{\pi}{12}, \qquad x_3 = \cos\frac{13\pi}{12} + i\sin\frac{13\pi}{12},$$

$$x_1 = \cos\frac{5\pi}{12} + i\sin\frac{5\pi}{12}, \qquad x_4 = \cos\frac{17\pi}{12} + i\sin\frac{17\pi}{12},$$

$$x_2 = \cos\frac{9\pi}{12} + i\sin\frac{9\pi}{12}, \qquad x_5 = \cos\frac{21\pi}{12} + i\sin\frac{21\pi}{12}.$$

Корню x_0 на плоскости P будет соответствовать вектор $\overrightarrow{Ox_0}$ с длиною, равною единице, причём этот вектор образует с положительным направлением оси $O\alpha$ угол, равный $\frac{\pi}{12}$, т. е. в 15°. Чтобы получить вектор $\overrightarrow{Ox_1}$, соответствующий корню x_1, надо повернуть вектор $\overrightarrow{Ox_0}$ против часовой стрелки на угол, равный $\frac{5\pi}{12} - \frac{\pi}{12} = \frac{\pi}{3}$, т. е. на 60°. На тот же угол против часовой стрелки придётся, очевидно, повернуть и вектор $\overrightarrow{Ox_1}$ для получения вектора $\overrightarrow{Ox_2}$, изображающего корень x_2, и т. д. Одним словом, концы векторов $\overrightarrow{Ox_0}$, $\overrightarrow{Ox_1}$, $\overrightarrow{Ox_2}$, $\overrightarrow{Ox_3}$, $\overrightarrow{Ox_4}$ и $\overrightarrow{Ox_5}$ будут находиться в вершинах правильного шестиугольника радиуса $R=1$.

Двучленное уравнение (2) можно свести к ещё более простому уравнению

$$x^n - 1 = 0. \tag{4}$$

Дело в том, что между корнями уравнений (2) и (4) имеет место следующая зависимость: *умножая один из корней n-й степени из a на всевозможные корни той же степени из единицы, мы получим все корни n-й степени из a.*

Доказательство. Будем обозначать, как и выше, корни n-й степени из a через x_k, а корни n-й степени из единицы — через α_k. Возьмём один из корней n-й степени из a, например x_0, и покажем, прежде всего, что $x_0\alpha_k$ есть также корень n-й степени из a. Для этой цели возведём $x_0\alpha_k$ в n-ю степень. Получим:

$$(x_0\alpha_k)^n = x_0^n\alpha_k^n = a \cdot 1 = a,$$

т. е. $x_0\alpha_k$ действительно оказалось корнем n-й степени из a: $x_0\alpha_k = x_k$.

Теперь пусть x_k — произвольный корень n-й степени из a. Рассмотрим частное $\frac{x_k}{x_0}$. Легко видеть, что $\frac{x_k}{x_0}$ есть корень n-й степени из единицы. В самом деле,

$$\left(\frac{x_k}{x_0}\right)^n = \frac{x_k^n}{x_0^n} = \frac{a}{a} = 1.$$

Таким образом, мы можем написать, что $\frac{x_k}{x_0} = \alpha_k$, откуда

$$x_k = x_0\alpha_k. \tag{5}$$

Итак, все корни n-й степени из a получаются по формуле (5), т. е. путём умножения корня x_0 n-й степени из a на всевозможные корни α_k n-й степени из единицы.

Среди корней n-й степени из единицы весьма важную роль в теории двучленных уравнений играют так называемые *первообразные корни.*

Корень ε n-й степени из единицы называется первообразным, если ε при возведении в степени 0, 1, 2, ..., n даёт *все* корни n-й степени из единицы.

Нетрудно убедиться, что *для всякого целого положительного числа n существует по меньшей мере один первообразный корень n-й степени из единицы.*

Чтобы убедиться в справедливости этого утверждения, обратимся к формуле извлечения корня n-й степени из комплексного числа. Согласно этой формуле получаем

$$\sqrt[n]{1} = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n} \qquad (k = 0,\ 1,\ 2,\ \ldots,\ n-1).$$

Отсюда по формуле Муавра

$$\sqrt[n]{1} = \varepsilon^k \qquad (k = 0,\ 1,\ 2,\ \ldots,\ n-1),$$

где

$$\varepsilon = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

Мы видим отсюда, что ε есть первообразный корень n-й степени из единицы.

§ 8. Уравнения второй и третьей степеней

Теперь перейдём к уравнениям второй и третьей степеней. Предварительно, однако, сделаем одно замечание, относящееся к алгебраическому уравнению

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0 \qquad (a_0 \neq 0) \tag{1}$$

любой степени. А именно, покажем, что соответствующей заменой неизвестного x новым неизвестным y можно добиться некоторого упрощения уравнения (1) — можно сделать равным нулю член с $(n-1)$-й степенью неизвестного. Для этой цели положим $x = y + \alpha$, где α — пока совершенно произвольное число. Получим:

$$a_0(y+\alpha)^n + a_1(y+\alpha)^{n-1} + \ldots + a_n = 0,$$

или, раскрывая скобки по биному Ньютона и располагая члены в порядке убывания степеней y:

$$a_0y^n + (na_0\alpha + a_1)y^{n-1} + \ldots + (a_0\alpha^n + a_1\alpha^{n-1} + \ldots + a_n) = 0.$$

Таким образом, если мы стремимся, чтобы в последнем уравнении исчез член с y^{n-1}, то надо, пользуясь произволом α, положить $na_0\alpha + a_1 = 0$. Отсюда получаем, что $\alpha = -\frac{a_1}{na_0}$.

Итак, полагая $x = y - \frac{a_1}{na_0}$, мы получим уравнение с неизвестным y, в котором член с y^{n-1} равен нулю.

Посмотрим, что даёт это преобразование для уравнения второй степени

$$ax^2 + bx + c = 0 \qquad (a \neq 0). \tag{2}$$

Здесь $a_0 = a$, $a_1 = b$, $n = 2$. Поэтому в соответствии с изложенным полагаем: $x = y - \frac{b}{2a}$. Получим:

$$a\left(y - \frac{b}{2a}\right)^2 + b\left(y - \frac{b}{2a}\right) + c = 0,$$

или после очевидных упрощений:

$$ay^2 - \frac{b^2 - 4ac}{4a} = 0,$$

откуда

$$y = \frac{\sqrt{b^2 - 4ac}}{2a}$$

и

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

Мы пришли к известной формуле решения квадратного уравнения. Обычно эту формулу записывают в виде

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{3}$$

подразумевая под $\sqrt{b^2 - 4ac}$ одно какое-нибудь значение корня квадратного из $b^2 - 4ac$. Формула (3), таким образом, даёт оба корня уравнения (2).

Остановимся несколько подробнее на извлечении корня квадратного из комплексного числа, поскольку решение уравнения второй степени связано с этой операцией.

На основании общей формулы извлечения корня n-й степени из комплексного числа мы можем в данном случае написать, что

$$\sqrt{A} = |\sqrt{r}|\left[\cos\left(\frac{\varphi}{2} + k\pi\right) + i\sin\left(\frac{\varphi}{2} + k\pi\right)\right] \qquad (k = 0,\ 1), \tag{4}$$

где φ — аргумент A и r — модуль A. Формулу (4) удобнее, однако, преобразовать в более простую формулу, избавляющую от необходимости находить модуль и аргумент комплексного числа A, когда A задано в двучленном виде $A = \alpha + \beta i$. Именно, выразим r, $\cos\varphi$ и $\sin\varphi$ через действительную часть α и коэффициент β при мнимой части комплексного числа A. Имеем:

$$r = |\sqrt{\alpha^2 + \beta^2}|, \quad \cos\varphi = \frac{\alpha}{|\sqrt{\alpha^2 + \beta^2}|}, \quad \sin\varphi = \frac{\beta}{|\sqrt{\alpha^2 + \beta^2}|};$$

символ абсолютной величины означает, что для модуля r числа A берётся положительное значение корня.

Отсюда

$$\cos\frac{\varphi}{2} = \pm\left|\sqrt{\frac{1 + \cos\varphi}{2}}\right| = \pm\left|\sqrt{\frac{\sqrt{\alpha^2 + \beta^2} + \alpha}{2}}\right|,$$

$$\sin\frac{\varphi}{2} = \pm\left|\sqrt{\frac{1 - \cos\varphi}{2}}\right| = \pm\left|\sqrt{\frac{\sqrt{\alpha^2 + \beta^2} - \alpha}{2}}\right|.$$

Знаки $\pm$ для $\cos\frac{\varphi}{2}$ и $\sin\frac{\varphi}{2}$ следует брать так, чтобы не было противоречия с соотношением

$$\cos\frac{\varphi}{2}\sin\frac{\varphi}{2}=\frac{1}{2}\sin\varphi=\frac{1}{2}\cdot\frac{\beta}{|\sqrt{\alpha^2+\beta^2}|}.$$

А именно, в случае $\beta>0$ числа $\cos\frac{\varphi}{2}$ и $\sin\frac{\varphi}{2}$ должны иметь одинаковые знаки, а при $\beta<0$ — противоположные знаки.

Принимая во внимание, что

$$\cos\left(\frac{\varphi}{2}+\pi\right)=-\cos\frac{\varphi}{2},\quad \sin\left(\frac{\varphi}{2}+\pi\right)=-\sin\frac{\varphi}{2},$$

получаем вместо формулы (4) следующую формулу:

$$\sqrt{A}=\sqrt{\alpha+\beta i}=\pm\left(\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}+\alpha}{2}}\right|\pm i\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}-\alpha}{2}}\right|\right). \qquad (5)$$

При $\beta\geqslant 0$ следует брать перед мнимой единицей знак плюс, а при $\beta<0$ — знак минус.

Пример 1. Найти, чему равен $\sqrt{21+20i}$.

Здесь $\alpha=21$, $\beta=20>0$. Поэтому, руководствуясь формулой (5), пишем:

$$\sqrt{21+20i}=\pm\left(\left|\sqrt{\frac{\sqrt{21^2+20^2}+21}{2}}\right|+i\left|\sqrt{\frac{\sqrt{21^2+20^2}-21}{2}}\right|\right)=$$
$$=\pm\left(|\sqrt{25}|+i|\sqrt{4}|\right)=\pm(5+2i).$$

Пример 2. Решить квадратное уравнение

$$(1+i)x^2-(7+3i)x+(22+20i)=0.$$

По формуле решения квадратного уравнения имеем:

$$x=\frac{(7+3i)+\sqrt{(7+3i)^2-4(1+i)(22+20i)}}{2(1+i)}$$

или

$$x=\frac{(7+3i)+\sqrt{32-126i}}{2(1+i)}.$$

Мы написали перед радикалом знак плюс, так как у нас $\sqrt{32-126i}$ означает любое значение квадратного корня.

Найдём теперь, чему равен $\sqrt{32-126i}$:

$$\sqrt{32-126i}=\pm\left(\left|\sqrt{\frac{\sqrt{32^2+126^2}+32}{2}}\right|-i\left|\sqrt{\frac{\sqrt{32^2+126^2}-32}{2}}\right|\right)=$$
$$=\pm(9-7i).$$

Отсюда

$$x=\frac{(7+3i)\pm(9-7i)}{2(1+i)},$$

$$x_1=\frac{(7+3i)+(9-7i)}{2(1+i)}=\frac{8-2i}{1+i}=3-5i,$$

$$x_2=\frac{(7+3i)-(9-7i)}{2(1+i)}=\frac{-1+5i}{1+i}=2+3i.$$

Займёмся исследованием квадратного уравнения с комплексными коэффициентами, т. е. выясним, при каких условиях уравнение имеет различные корни и при каких условиях оно имеет кратные корни. Нам придётся воспользоваться так называемыми формулами Вьета. Мы их выведем для общего случая уравнения n-й степени, так как эти формулы понадобятся и для уравнений выше второй степени.

Пусть

$$x^n+a_1x^{n-1}+\ldots+a_n=0$$

— алгебраическое уравнение с комплексными коэффициентами $a_1, \ldots, a_n$ и со старшим коэффициентом, равным единице. Обозначим его корни через $x_1, x_2, \ldots, x_n$. Тогда можно будет написать следующее разложение многочлена $f(x)=x^n+a_1x^{n-1}+\ldots+a_n$ на линейные множители:

$$x^n+a_1x^{n-1}+\ldots+a_n=(x-x_1)(x-x_2)\ldots(x-x_n).$$

Отсюда, производя перемножение линейных множителей, получаем:

$$x^n+a_1x^{n-1}+\ldots+a_n=$$
$$=x^n-(x_1+x_2+\ldots+x_n)x^{n-1}+\ldots+(-1)^n x_1x_2\ldots x_n.$$

Но если два многочлена равны, то их коэффициенты при одинаковых степенях x должны совпадать. Следовательно,

$$\left.\begin{aligned} a_1&=-(x_1+x_2+\ldots+x_n),\\ a_2&=+(x_1x_2+x_1x_3+\ldots+x_{n-1}x_n),\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ a_n&=(-1)^n x_1x_2\ldots x_n. \end{aligned}\right\} \qquad (A)$$

В правых частях соотношений (А) стоят всевозможные произведения корней по одному, по два, по три и т. д. Мы пришли к формулам Вьета.

Вернёмся к квадратному уравнению. Для большей простоты можно всегда предположить, что старший коэффициент уравнения равен единице; в противном случае мы разделили бы обе части уравнения на старший коэффициент. В этом предположении квадратное уравнение будет выглядеть так:

$$x^2+px+q=0. \qquad (6)$$

Обозначим корни уравнения (6) через x_1 и x_2. Тогда по формулам Вьета имеем:

$$p = -(x_1 + x_2), \quad q = x_1 x_2. \tag{7}$$

Корни x_1 и x_2 уравнения (6) найдутся по формуле

$$x = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}.$$

Мы пишем перед радикалом знак плюс, так как у нас $\sqrt{\frac{p^2}{4} - q}$ означает любое значение квадратного корня.

Подкоренное выражение $\frac{p^2}{4} - q$ обычно называется *дискриминантом* квадратного уравнения. Исходя из формул Вьета, легко показать, что *уравнение* (6) *тогда и только тогда имеет двукратный корень, когда его дискриминант равен нулю.*

В самом деле, если уравнение (6) имеет один двукратный корень, то это означает, что уравнение имеет два равных корня: $x_1 = x_2$. Отсюда $p = -(x_1 + x_2) = -2x_1$, $q = x_1 x_2 = x_1^2$ и

$$\frac{p^2}{4} - q = \frac{(-2x_1)^2}{4} - x_1^2 = x_1^2 - x_1^2 = 0.$$

Обратно, если дискриминант $\frac{p^2}{4} - q$ равен нулю, то, подставляя вместо p и q их выражения через корни x_1 и x_2 по формулам Вьета, получаем:

$$\frac{p^2}{4} - q = \frac{(x_1 + x_2)^2}{4} - x_1 x_2 = \frac{(x_1 - x_2)^2}{4} = 0,$$

откуда $(x_1 - x_2)^2 = 0$, или $x_1 = x_2$.

Если, в частности, коэффициенты уравнения (6) действительны, то с помощью дискриминанта можно установить, когда корни уравнения действительны и когда они являются мнимыми. Очевидно, что корни уравнения (6) будут действительными в том и только в том случае, когда его дискриминант $\frac{p^2}{4} - q$ неотрицателен, так как в этом случае корень квадратный из $\frac{p^2}{4} - q$ имеет действительные значения.

В дополнение к только что сказанному можно отметить, что 1) уравнение (6) имеет различные действительные отрицательные корни, если $\frac{p^2}{4} - q > 0$ и коэффициенты p и q положительны; 2) уравнение (6) имеет различные действительные положительные корни, если $\frac{p^2}{4} - q > 0$, коэффициент p отрицателен, а коэффициент q положителен; 3) уравнение (6) имеет один корень положительный,

а другой отрицательный, если $\frac{p^2}{4} - q > 0$ и коэффициент q отрицателен; 4) уравнение (6) имеет при $\frac{p^2}{4} - q = 0$ только один двукратный действительный корень, положительный или отрицательный в зависимости от того, будет ли p отрицательно или положительно.

В справедливости всех этих утверждений легко убедиться с помощью формул Вьета.

Перейдём к уравнению третьей степени:

$$x^3 + ax^2 + bx + c = 0.$$

Без ограничения общности старший коэффициент можно предполагать равным единице. Подвергнем это уравнение уже знакомому упрощению — сделаем член с квадратом неизвестного равным нулю, для чего положим $x = y - \frac{a}{3}$. Получим в результате так называемое *каноническое уравнение третьей степени*:

$$y^3 + py + q = 0, \tag{8}$$

где $p = b - \frac{1}{3}a^2$, $q = \frac{2a^3}{27} - \frac{ab}{3} + c$.

Чтобы найти корни уравнения (8), воспользуемся следующим приёмом: подразумевая под y любой из корней уравнения (8), рассмотрим квадратное уравнение

$$z^2 - yz - \frac{p}{3} = 0.$$

Если u, v — корни этого квадратного уравнения, то по формулам Вьета

$$y = u + v \text{ и } uv = -\frac{p}{3}. \tag{9}$$

Отсюда, подставляя выражение y через u, v в уравнение (8), имеем:

$$(u+v)^3 + p(u+v) + q = 0.$$

Раскроем скобки и произведём некоторую перегруппировку членов. Получим тогда, что

$$(u+v)(3uv+p) + (u^3 + v^3 + q) = 0.$$

Но в силу равенства (9) $3uv + p = 0$. Следовательно,

$$u^3 + v^3 + q = 0,$$

или

$$u^3 + v^3 = -q. \tag{10}$$

Кроме того, возводя обе части равенства (9) в куб, получаем:

$$u^3v^3 = -\left(\frac{p}{3}\right)^3. \tag{11}$$

Из равенств (10) и (11) видно на основании формул Вьета, что u^3 и v^3 являются корнями следующего квадратного уравнения:

$$t^2 + qt - \left(\frac{p}{3}\right)^3 = 0.$$

Решаем это квадратное уравнение:

$$t_1 = u^3 = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3},$$

$$t_2 = v^3 = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3},$$

откуда

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

$$v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

и

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \tag{I}$$

Мы пришли к формуле Кардана решения уравнения третьей степени. При вычислении корней уравнения (8) третьей степени следует, однако, иметь в виду, что u и v не независимы, а связаны друг с другом равенством

$$uv = -\frac{p}{3}.$$

Пример 3. Решить с помощью формулы Кардана уравнение

$$x^3 + 15x + 124 = 0.$$

Это уравнение является каноническим; поэтому можно сразу обратиться к формуле Кардана.

Здесь $p = 15$, $q = 124$. Следовательно, в данном случае

$$u = \sqrt[3]{-62 + \sqrt{3969}} = \sqrt[3]{-62 + 63} = \sqrt[3]{1}.$$

Находим все три значения u; так как

$$u = \sqrt[3]{1} = \sqrt[3]{\cos 0 + i \sin 0} = \cos\frac{2\pi k}{3} + i \sin\frac{2\pi k}{3}, \qquad (k = 0,\ 1,\ 2)$$

то

$$u_0 = \cos 0 + i \sin 0 = 1,$$

$$u_1 = \cos \frac{2\pi}{3} + i \sin \frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2},$$

$$u_2 = \cos \frac{4\pi}{3} + i \sin \frac{4\pi}{3} = -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Соответствующие значения v мы найдём, пользуясь равенством $uv = -\frac{p}{3} = -5$. Имеем:

$$v_0 = -\frac{5}{1} = -5,$$

$$v_1 = -\frac{5}{-\frac{1}{2} + i\frac{\sqrt{3}}{2}} = \frac{5}{2} + i\frac{5\sqrt{3}}{2},$$

$$v_2 = -\frac{5}{-\frac{1}{2} - i\frac{\sqrt{3}}{2}} = \frac{5}{2} - i\frac{5\sqrt{3}}{2}.$$

Отсюда без труда находим все три корня уравнения:

$$x_0 = u_0 + v_0 = -4,$$
$$x_1 = u_1 + v_1 = 2 + 3i\sqrt{3},$$
$$x_2 = u_2 + v_2 = 2 - 3i\sqrt{3}.$$

Пример 4. Решить с помощью формулы Кардана уравнение

$$x^3 + 3x^2 - 6x + 3 = 0.$$

Приводим это уравнение к каноническому виду, полагая $x = y - 1$. Получаем:

$$y^3 - 9y + 11 = 0.$$

Отсюда

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} = \sqrt[3]{-\frac{11}{2} + \sqrt{\frac{121}{4} - 27}} =$$

$$= \sqrt[3]{\frac{-11 + \sqrt{13}}{2}}.$$

Здесь удобнее сначала найти действительное значение u_0 корня третьей степени из $\frac{-11 + \sqrt{13}}{2}$. С точностью до четвёртого знака после запятой находим, что

$$u_0 \approx -1{,}5463.$$

Остальные два значения u мы найдём, умножая u_0 на корни третьей степени из единицы:

$$\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}=-\frac{1}{2}+i\frac{\sqrt{3}}{2},$$

$$\cos\frac{4\pi}{3}+i\sin\frac{4\pi}{3}=-\frac{1}{2}-i\frac{\sqrt{3}}{2}.$$

Получаем:

$$u_1=\frac{u_0}{2}(-1+i\sqrt{3}),\quad u_2=\frac{u_0}{2}(-1-i\sqrt{3}).$$

Соответствующие значения v найдутся из равенства $uv=-\frac{p}{3}=3$:

$$v_0=\frac{3}{u_0}\approx-1{,}9401,$$

$$v_1=\frac{3}{u_1}=-\frac{1}{2}v_0(1+i\sqrt{3})\approx 0{,}9700\,(1+i\sqrt{3}),$$

$$v_2=\frac{3}{u_2}=-\frac{1}{2}v_0(1-i\sqrt{3})\approx 0{,}9700\,(1-i\sqrt{3}).$$

Таким образом,

$$y_0=u_0+v_0\approx-3{,}4864,$$
$$y_1=u_1+v_1\approx 1{,}7431+0{,}1969\,i\sqrt{3},$$
$$y_2=u_2+v_2\approx 1{,}7431-0{,}1969\,i\sqrt{3},$$

или

$$y_0\approx-3{,}4864,$$
$$y_1\approx 1{,}7431+0{,}3410\,i,$$
$$y_2\approx 1{,}7431-0{,}3410\,i.$$

Вспомнив теперь, что $x=y-1$, получаем для корней данного уравнения следующие значения:

$$x_0=y_0-1\approx-4{,}4864,$$
$$x_1=y_1-1\approx 0{,}7431+0{,}3410\,i,$$
$$x_2=y_2-1\approx 0{,}7431-0{,}3410\,i.$$

Пусть u_0 — одно из значений u. Обозначим через ε первообразный корень третьей степени из единицы. Тогда остальные два значения u могут быть записаны следующим образом:

$$u_1=u_0\varepsilon,\quad u_2=u_0\varepsilon^2.$$

Отсюда получаем для соответствующих значений v:

$$v_0=-\frac{p}{3u_0},\quad v_1=-\frac{p}{3u_0\varepsilon}=-\frac{p}{3u_0\varepsilon^3}\cdot\varepsilon^2=-\frac{p}{3u_0}\varepsilon^2=v_0\varepsilon^2,$$
$$v_2=-\frac{p}{3u_0\varepsilon^2}=-\frac{p}{3u_0\varepsilon^3}\cdot\varepsilon=-\frac{p}{3u_0}\varepsilon=v_0\varepsilon.$$

Таким образом, корни уравнения (8) можно находить по формулам

$$\left.\begin{aligned} y_0 &= u_0 + v_0, \\ y_1 &= u_0\varepsilon + v_0\varepsilon^2, \\ y_2 &= u_0\varepsilon^2 + v_0\varepsilon, \end{aligned}\right\} \tag{12}$$

где u_0 — одно из значений u (какое именно, безразлично), $v_0 = -\frac{p}{3u_0}$, а ε — какой-нибудь первообразный корень третьей степени из единицы.

Если в качестве ε взять $\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$, то формулы (12) принимают ещё более удобный для вычисления вид:

$$\left.\begin{aligned} y_0 &= u_0 + v_0, \\ y_1 &= -\frac{u_0 + v_0}{2} + i\frac{(u_0 - v_0)\sqrt{3}}{2}, \\ y_2 &= -\frac{u_0 + v_0}{2} - i\frac{(u_0 - v_0)\sqrt{3}}{2}. \end{aligned}\right\} \tag{12*}$$

Выражение $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$, фигурирующее под корнем квадратным в формуле Кардана, иногда называют *дискриминантом* уравнения (8) [1]. Посмотрим, что получается в случае $\Delta = 0$ и $\Delta \neq 0$.

Если $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$, то $u = \sqrt[3]{-\frac{q}{2}}$. Это выражение u можно несколько упростить. А именно,

$$u = \sqrt[3]{\frac{\left(-\frac{q}{2}\right)^3}{\left(\frac{q}{2}\right)^2}} = -\frac{q}{2\sqrt[3]{\left(\frac{q}{2}\right)^2}}.$$

Так как $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$, то $\left(\frac{q}{2}\right)^2 = -\left(\frac{p}{3}\right)^3$. Следовательно,

$$u = -\frac{q}{2\sqrt[3]{-\left(\frac{p}{3}\right)^3}},$$

откуда в качестве одного из значений u получается следующее выражение:

$$u_0 = -\frac{q}{2\left(-\frac{p}{3}\right)} = \frac{3q}{2p}.$$

---

[1]) Более употребительно называть дискриминантом не Δ, а $-\Delta$.

Соответствующее значение v_0 будет равно

$$v_0 = -\frac{p}{3u_0} = -\frac{2p^2}{9q} = \frac{6\cdot\left[-\left(\frac{p}{3}\right)^3\right]}{pq} = -\frac{6\left(\frac{q}{2}\right)^2}{pq} = \frac{3q}{2p} = u_0.$$

Обращаясь к формулам (12*), мы, таким образом, получаем:

$$y_0 = u_0 + u_0 = 2u_0 = \frac{3q}{p}.$$
$$y_1 = -\frac{2u_0}{2} = -u_0 = -\frac{3q}{2p},$$
$$y_2 = -\frac{2u_0}{2} = -u_0 = -\frac{3q}{2p}.$$

Итак, *если* $\Delta = 0$, *то уравнение* (8) *при* $p \neq 0$ *и* $q \neq 0$ *имеет один простой корень* y_0 *и один двукратный корень* $y_1 = y_2$. *Эти корни можно найти, не прибегая к извлечению корней второй и третьей степеней, а именно, их можно вычислить по формулам*

$$y_0 = \frac{3q}{p}, \quad y_1 = y_2 = -\frac{3q}{2p}. \tag{13}$$

Пример 5. Решить уравнение

$$x^3 - 3x^2 - 9x + 27 = 0.$$

Приводим, прежде всего, это уравнение к каноническому виду, для чего полагаем $x = y + 1$. Получаем:

$$y^3 - 12y + 16 = 0.$$

Легко видеть, что здесь $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$, и мы можем воспользоваться формулами (13):

$$y_0 = \frac{3q}{p} = \frac{48}{-12} = -4, \quad y_1 = y_2 = -\frac{48}{-24} = 2.$$

Обратимся к случаю $\Delta \neq 0$ и покажем, что *если* $\Delta \neq 0$, *то уравнение* (8) *имеет три различных корня.*

Доказательство. Предположим противное: пусть уравнение (8) имеет два корня, равных одному и тому же числу α; третий же корень пусть равен β. Тогда по формулам Вьета получаем, что

$$2\alpha + \beta = 0, \quad \alpha^2 + 2\alpha\beta = p, \quad -\alpha^2\beta = q.$$

Следовательно, $\beta = -2\alpha$ и

$$p = -3\alpha^2, \quad q = 2\alpha^3.$$

Таким образом,

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = \alpha^6 - \alpha^6 = 0,$$

что противоречит условию $\Delta \neq 0$.

До сих пор мы предполагали коэффициенты кубического уравнения любыми комплексными числами. Рассмотрим теперь наиболее часто встречающийся случай кубического уравнения с действительными коэффициентами. Мы увидим, что и в этом случае дискриминант Δ играет существенную роль в исследовании кубического уравнения.

А) $\Delta > 0$. Так как здесь $\Delta \neq 0$, то все три корня уравнения (8) должны быть различными. Выясним, сколько среди них будет действительных корней.

Обращаясь к выражению

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}},$$

легко усмотреть, что под кубическим корнем находится действительное число, так как $\Delta > 0$. Следовательно, одно из значений u должно быть действительным. Примем его за u_0. Тогда v_0 будет также действительным. Отсюда на основании формул (12*) заключаем, что уравнение (8) имеет только один действительный корень, а именно $y_0 = u_0 + v_0$. Выясним, когда этот корень положителен и когда отрицателен.

Пусть $p > 0$. Тогда

$$\left|-\frac{q}{2}\right| < \left|\sqrt{\Delta}\right|,$$

и u_0 должно быть положительным, а число v_0, равное действительному значению

$$\sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}},$$

должно быть, очевидно, отрицательным. Далее, при $q > 0$

$$\left|-\frac{q}{2} + \sqrt{\Delta}\right| < \left|-\frac{q}{2} - \sqrt{\Delta}\right|,$$

а при $q < 0$

$$\left|-\frac{q}{2} + \sqrt{\Delta}\right| > \left|-\frac{q}{2} - \sqrt{\Delta}\right|.$$

Таким образом, если $q > 0$, то $|u_0| \leqslant |v_0|$, вследствие чего $y_0 = u_0 + v_0$ будет отрицательным; если же $q < 0$, то $|u_0| > |v_0|$ и потому y_0 будет положительным.

Пусть теперь $p < 0$. Тогда

$$\left|-\frac{q}{2}\right| > \left|\sqrt{\Delta}\right|$$

и u_0 должно быть отрицательным при $q > 0$ и положительным при $q < 0$, а число v_0, равное действительному значению

$$\sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}},$$

будет также при $q>0$ отрицательным и при $q<0$ положительным. Отсюда корень $y_0=u_0+v_0$ при $q>0$ будет отрицателен, а при $q<0$ положителен.

Итак, *если* $\Delta>0$, *то уравнение* (8) *имеет лишь один действительный корень, причём при* $q>0$ *этот корень будет отрицательным, а при* $q<0$ *— положительным.*

Б) $\Delta=0$. Мы знаем, что при $\Delta=0$, $p\neq 0$, $q\neq 0$ уравнение имеет два равных корня. Учитывая, что теперь рассматривается уравнение с действительными коэффициентами, можно сделать следующее заключение: *при* $\Delta=0$, $p\neq 0$, $q\neq 0$ *все три корня уравнения* (8) *действительны, причём два из них равны; иными словами, уравнение имеет один простой действительный корень и один двукратный действительный корень:*

$$y_0=\frac{3q}{p},\quad y_1=y_2=-\frac{3q}{2p}.$$

Из $\Delta=\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3=0$ вытекает, что $\left(\frac{p}{3}\right)^3=-\left(\frac{q}{2}\right)^2<0$, откуда $p<0$. Таким образом, *если* $q>0$, *то простой корень* y_0 *будет отрицательным, а двукратный корень* $y_1=y_2$ *будет положительным; если же* $q<0$, *то* y_0 *будет положительно, а* $y_1=y_2$ *— отрицательно.*

В) $\Delta<0$. Этот случай известен под названием *неприводимого случая* и является в следующем отношении примечательным. А именно, так как корень третьей степени здесь приходится извлекать из мнимых чисел, то u и v являются мнимыми. И тем не менее все три корня будут действительны. В самом деле, так как $\Delta<0$, то мы можем положить $\Delta=-\alpha^2$, где α — некоторое действительное положительное число. Тогда

$$u=\sqrt[3]{-\frac{q}{2}+\alpha i}.$$

Найдём модуль r и аргумент φ подкоренного выражения. Имеем:

$$r=\left|\sqrt{\left(-\frac{q}{2}\right)^2+\alpha^2}\right|=\left|\sqrt{\frac{q^2}{4}-\Delta}\right|=$$

$$=\left|\sqrt{\frac{q^2}{4}-\frac{q^2}{4}-\frac{p^3}{27}}\right|=\left|\sqrt{-\frac{p^3}{27}}\right|,$$

$$\cos\varphi=-\frac{q}{2r},\quad \sin\varphi=\frac{\alpha}{r}>0.$$

Таким образом,

$$u=\sqrt[3]{r(\cos\varphi+i\sin\varphi)}=\left|\sqrt[3]{r}\right|\left(\cos\frac{\varphi+2k\pi}{3}+i\sin\frac{\varphi+2k\pi}{3}\right).$$

Полагая последовательно $k=0$, 1, 2, получаем все три значения u:

$$u_0=|\sqrt[3]{r}|\left(\cos\frac{\varphi}{3}+i\sin\frac{\varphi}{3}\right),$$

$$u_1=|\sqrt[3]{r}|\left(\cos\frac{\varphi+2\pi}{3}+i\sin\frac{\varphi+2\pi}{3}\right),$$

$$u_2=|\sqrt[3]{r}|\left(\cos\frac{\varphi+4\pi}{3}+i\sin\frac{\varphi+4\pi}{3}\right).$$

Известно, что произведение комплексного числа z на сопряжённое комплексное число $\bar{z}$ равно квадрату модуля z:

$$z\bar{z}=|z|^2.$$

Руководствуясь этим, мы можем легко определить v_0, v_1, v_2. Для этой цели обратимся к выражению

$$u=|\sqrt[3]{r}|\left(\cos\frac{\varphi+2k\pi}{3}+i\sin\frac{\varphi+2k\pi}{3}\right).$$

Мы видим, что модуль u равен

$$|\sqrt[3]{r}|=\left|\sqrt[3]{\sqrt{-\frac{p^3}{27}}}\right|=\left|\sqrt{-\frac{p}{3}}\right|.$$

Отсюда квадрат модуля u будет равен $-\frac{p}{3}$. Следовательно, $u\bar{u}=-\frac{p}{3}$. Но u и v связаны тем же самым соотношением: $uv=-\frac{p}{3}$. Значит, $v=\bar{u}$, и мы получаем, что

$$v_0=\bar{u}_0=|\sqrt[3]{r}|\left(\cos\frac{\varphi}{3}-i\sin\frac{\varphi}{3}\right),$$

$$v_1=\bar{u}_1=|\sqrt[3]{r}|\left(\cos\frac{\varphi+2\pi}{3}-i\sin\frac{\varphi+2\pi}{3}\right),$$

$$v_2=\bar{u}_2=|\sqrt[3]{r}|\left(\cos\frac{\varphi+4\pi}{3}-i\sin\frac{\varphi+4\pi}{3}\right).$$

Теперь все три корня уравнения (8) найдутся без труда:

$$\left.\begin{aligned} y_0&=u_0+v_0=2|\sqrt[3]{r}|\cos\frac{\varphi}{3},\\ y_1&=u_1+v_1=2|\sqrt[3]{r}|\cos\frac{\varphi+2\pi}{3},\\ y_2&=u_2+v_2=2|\sqrt[3]{r}|\cos\frac{\varphi+4\pi}{3}. \end{aligned}\right\} \qquad (14)$$

Из формул (14) видно, что корни y_0, y_1, y_2 действительны и различны. Кроме того, исходя из формул (14), можно показать,

что при $q > 0$ уравнение (8) имеет два положительных корня, а при $q < 0$ — только один положительный корень.

В самом деле, если $q > 0$, то $\cos\varphi < 0$. Так как $\sin\varphi > 0$, то угол φ должен лежать во второй четверти. Отсюда $\frac{\varphi}{3} > \frac{\pi}{6}$ и лежит в первой четверти, а $\frac{\varphi + 4\pi}{3}$ — в четвёртой, вследствие чего корни y_0 и y_2 положительны. Если $q < 0$, то аналогичным образом убеждаемся, что только y_0 будет положительным.

Итак, *в случае* $\Delta < 0$ *уравнение* (8) *имеет три действительных различных корня, причём при* $q > 0$ *два корня положительны, а при* $q < 0$ *только один корень положителен.*

Корни y_0, y_1, y_2 вычисляются по формулам (14) довольно легко, если пользоваться таблицами логарифмов значений тригонометрических функций.

Формула (I) на стр. 211 обладает тем недостатком, что она в случае отрицательного Δ выражает действительные корни уравнения (8) с действительными коэффициентами в мнимом виде. Для Кардана и его современников случай отрицательного Δ казался парадоксальным, так как в то время понятие комплексного числа ещё не имело конкретного истолкования и операции извлечения квадратного корня из отрицательных чисел и извлечения кубического корня из комплексных чисел считались невозможными. Для математиков того времени было удивительным то обстоятельство, что в случае $\Delta < 0$ получались с помощью этих невозможных операций действительные числа. Были предприняты многочисленные попытки освободиться от мнимостей в формуле Кардана, но эти попытки кончались неудачей. С помощью рассуждений, выходящих за пределы нашей статьи, можно показать, что корни уравнения (8) с действительными коэффициентами в случае $\Delta < 0$ никаким способом нельзя выразить через радикалы с действительными подкоренными выражениями. В силу этого обстоятельства случай $\Delta < 0$ и получил наименование *неприводимого случая.*

Другой недостаток формулы (I) состоит в том, что она часто представляет рациональные корни в иррациональном виде. Приведём пример.

Пример 6. С помощью общего приёма нахождения рациональных корней многочленов с рациональными коэффициентами легко убедиться, что уравнение

$$x^3 - x - 6 = 0$$

имеет рациональный корень $x_0 = 2$. Так как для данного уравнения

$$\Delta = \frac{242}{27} > 0,$$

то 2 является единственным действительным корнем уравнения.

Посмотрим теперь, что даёт формула (I). Обозначим через u_0, v_0 действительные значения кубических корней в формуле (I):

$$u_0 = \sqrt[3]{3 + \frac{11}{9}\sqrt{6}}, \quad v_0 = \sqrt[3]{3 - \frac{11}{9}\sqrt{6}}.$$

Мы видим, что u_0 и v_0 являются иррациональными числами. Таким образом, формула (I) даёт для корня $x_0 = 2$ довольно сложное выражение

$$x_0 = \sqrt[3]{3 + \frac{11}{9}\sqrt{6}} + \sqrt[3]{3 - \frac{11}{9}\sqrt{6}}$$

(для каждого кубического корня берётся действительное значение), которое приходится вычислять приближённо, так что на практике получается число, весьма близкое к 2, но не 2.

Вследствие этого недостатка рекомендуется рациональные корни кубического уравнения с рациональными коэффициентами находить не по формуле (I), а с помощью общего приёма вычисления рациональных корней, изложенного в § 4.

§ 9. Уравнение четвёртой степени

Мы начнём с изложения самого раннего способа решения уравнения четвёртой степени

$$x^4 + ax^3 + bx^2 + cx + d = 0, \tag{1}$$

решения, принадлежащего Луиджи Феррари — ученику Кардана. Этот способ, впрочем, был нами использован ещё в § 5 для разложения многочлена четвёртой степени с рациональными коэффициентами на множители.

Мы будем предполагать, как и в предыдущем параграфе, коэффициенты уравнения комплексными (и, в частности, действительными) числами.

Подвергнем многочлен

$$f(x) = x^4 + ax^3 + bx^2 + cx + d,$$

находящийся в левой части уравнения (1), тем же самым преобразованиям, что и в § 5. В результате уравнение (1) превратится в

$$\left(x^2 + \frac{ax}{2} + y\right)^2 - (Ax^2 + Bx + C) = 0, \tag{2}$$

где

$$A = 2y + \frac{a^2}{4} - b, \quad B = ay - c, \quad C = y^2 - d.$$

Подберём теперь вспомогательную величину y так, чтобы квадратный трёхчлен $Ax^2 + Bx + C$ был полным квадратом некоторого линейного многочлена $\alpha x + \beta$. Мы знаем, что это может

быть в том и только в том случае, когда y является корнем кубической резольвенты

$$(ay-c)^2=4\left(2y+\frac{a^2}{4}-b\right)(y^2-d). \tag{3}$$

Пусть y_0 — один из корней уравнения (3). Тогда при $y=y_0$

$$Ax^2+Bx+C=(\alpha x+\beta)^2,$$

в результате чего уравнение (2) перепишется следующим образом:

$$\left(x^2+\frac{ax}{2}+y_0\right)^2-(\alpha x+\beta)^2=0,$$

или, разлагая разность квадратов в произведение суммы на разность:

$$\left[x^2+\left(\frac{a}{2}+\alpha\right)x+(y_0+\beta)\right]\cdot\left[x^2+\left(\frac{a}{2}-\alpha\right)x+(y_0-\beta)\right]=0.$$

Отсюда, решая квадратные уравнения

$$\left.\begin{aligned} x^2+\left(\frac{a}{2}+\alpha\right)x+(y_0+\beta)&=0,\\ x^2+\left(\frac{a}{2}-\alpha\right)x+(y_0-\beta)&=0,\end{aligned}\right\} \tag{4}$$

мы получим все четыре корня уравнения (1).

Итак, *решение уравнения* (1) *сводится к решению уравнения третьей степени — кубической резольвенты — и решению квадратных уравнений* (4).

При решении того или иного уравнения четвёртой степени мы рекомендуем проводить последовательно преобразования Феррари, а не пользоваться готовыми формулами. В качестве образца приводим решение следующего уравнения.

Пример 1. Решить с помощью способа Феррари уравнение

$$x^4+2x^3+5x^2+6x+9=0.$$

Прежде всего перенесём в правую часть уравнения с противоположными знаками все члены, степень которых не выше двух:

$$x^4+2x^3=-5x^2-6x-9$$

или

$$(x^2)^2+2x^2x=-5x^2-6x-9.$$

Если к обеим частям последнего уравнения прибавить x^2, то в левой части получится полный квадрат. Производим это преобразование:

$$(x^2)^2+2x^2x+x^2=-5x^2-6x-9+x^2$$

или

$$(x^2+x)^2=-4x^2-6x-9.$$

Теперь к обеим частям получившегося уравнения прибавляем $2(x^2+x)y+y^2$. От этого левая часть не перестанет быть полным квадратом:

$$(x^2+x)^2+2(x^2+x)y+y^2=-4x^2-6x-9+2(x^2+x)y+y^2,$$

или

$$(x^2+x+y)^2=(2y-4)x^2+(2y-6)x+(y^2-9). \qquad (5)$$

Возьмем теперь y таким, чтобы и правая часть уравнения (5) была полным квадратом. Для этого y должно быть корнем кубической резольвенты. Чтобы получить кубическую резольвенту, надо воспользоваться условием $B^2=4AC$. В данном случае $A=2y-4$, $B=2y-6$ и $C=y^2-9$. Следовательно,

$$(2y-6)^2=4(2y-4)(y^2-9)$$

или после некоторых упрощений:

$$(y-3)[(y-3)-(2y-4)(y+3)]=0.$$

Мы видим отсюда, что в качестве y_0 можно взять 3.

Возвращаемся к уравнению (5) и заменяем в нём y значением $y_0=3$:

$$(x^2+x+3)^2=2x^2$$

или

$$(x^2+x+3)^2-(\sqrt{2}x)^2=0,$$

$$[x^2+(1+\sqrt{2})x+3][x^2+(1-\sqrt{2})x+3]=0,$$

откуда

$$\left.\begin{aligned} x^2+(1+\sqrt{2})x+3&=0,\\ x^2+(1-\sqrt{2})x+3&=0.\end{aligned}\right\} \qquad (6)$$

Решая уравнения (6), получаем все четыре корня данного уравнения четвёртой степени, а именно,

$$x_{1,2}=-\frac{1+\sqrt{2}}{2}\pm i\frac{\sqrt{9-2\sqrt{2}}}{2},$$

$$x_{3,4}=\frac{\sqrt{2}-1}{2}\pm i\frac{\sqrt{9+2\sqrt{2}}}{2}.$$

Приведём ещё один способ решения уравнения четвёртой степени. Он принадлежит знаменитому петербургскому академику Леонарду Эйлеру и замечателен в том отношении, что непосредственно выражает корни уравнения четвёртой степени через корни кубической резольвенты.

Положим, прежде всего, $x = y - \frac{a}{4}$. Тогда полное уравнение четвёртой степени (1) превратится в четырёхчленное уравнение

$$y^4 + py^2 + qy + r = 0. \tag{7}$$

Наряду с уравнением (7) рассмотрим кубическое уравнение относительно z:

$$z^3 - 2yz^2 + mz + n = 0, \tag{8}$$

где y — любой из корней уравнения (7), а коэффициенты m и n пока произвольны.

Если корни уравнения (8) обозначить через u, v, w, то по формулам Вьета будем иметь

$$2y = u + v + w, \quad m = uv + uw + vw, \quad n = -uvw.$$

Возводим обе части равенства

$$2y = u + v + w \tag{9}$$

в квадрат:

$$4y^2 = u^2 + v^2 + w^2 + 2(uv + uw + vw). \tag{10}$$

Обе части равенства (10) возводим снова в квадрат:

$$16y^4 = (u^2 + v^2 + w^2)^2 + 4(uv + uw + vw)(u^2 + v^2 + w^2) + \\ + 4(u^2v^2 + u^2w^2 + v^2w^2) + 8uvw(u + v + w). \tag{11}$$

Подставляя в уравнение (7) вместо y, y^2, y^4 их выражения из равенств (9), (10) и (11), получаем после некоторых упрощений:

$$\left.\begin{array}{r}(u^2 + v^2 + w^2)^2 + 4(uv + uw + vw)(u^2 + v^2 + w^2 + 2p) + \\ + 4p(u^2 + v^2 + w^2) + 8(uvw + q)(u + v + w) + \\ + 4(u^2v^2 + u^2w^2 + v^2w^2) + 16r = 0.\end{array}\right\} \tag{12}$$

Подберём теперь коэффициенты m и n уравнения (8) так, чтобы уравнение (12) максимально упростилось. А именно, положим:

$$u^2 + v^2 + w^2 + 2p = 0, \quad uvw + q = 0^1).$$

Тогда уравнение (12) превратится в

$$u^2v^2 + u^2w^2 + v^2w^2 = p^2 - 4r.$$

---

[1]) Легко убедиться, что такой подбор коэффициентов вполне возможен. Прежде всего из равенства $uvw + q = 0$ следует, что $n = -uvw = q$. Затем, заменяя в правой части равенства (10) $u^2 + v^2 + w^2$ и $uv + uw + vw$ их значениями $-2p$ и m, получаем:

$$4y^2 = -2p + 2m,$$

откуда $m = 2y^2 + p$.

Отсюда следует, что u, v, w удовлетворяют следующей системе уравнений:

$$\left.\begin{aligned} u^2+v^2+w^2 &= -2p,\\ u^2v^2+u^2w^2+v^2w^2 &= p^2-4r,\\ u^2v^2w^2 &= q^2. \end{aligned}\right\} \tag{13}$$

Из равенств (13) следует на основании формул Вьета, что u^2, v^2, w^2 являются корнями следующего кубического уравнения

$$z^3+2pz^2+(p^2-4r)z-q^2=0, \tag{14}$$

которое переходит в уравнение кубической резольвенты для уравнения (7), если заменить здесь z через $2z'-p$.

Таким образом, решая уравнение (14), мы получим три его корня

$$z_1=u^2,\ z_2=v^2,\ z_3=w^2,$$

откуда

$$u=\sqrt{z_1},\quad v=\sqrt{z_2},\quad w=\sqrt{z_3}.$$

Значения радикалов $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$ следует выбирать с таким расчётом, чтобы выполнялось равенство

$$\sqrt{z_1}\cdot\sqrt{z_2}\cdot\sqrt{z_3}=-q. \tag{15}$$

Очевидно, что значения двух радикалов могут быть выбраны произвольно, а значение третьего радикала придётся уже брать, исходя из равенства (15).

Выбрав указанным образом значения радикалов $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$, мы получим все четыре корня уравнения (7) по формулам

$$\left.\begin{aligned} 2y_1 &= \sqrt{z_1}+\sqrt{z_2}+\sqrt{z_3},\\ 2y_2 &= \sqrt{z_1}-\sqrt{z_2}-\sqrt{z_3},\\ 2y_3 &= -\sqrt{z_1}+\sqrt{z_2}-\sqrt{z_3},\\ 2y_4 &= -\sqrt{z_1}-\sqrt{z_2}+\sqrt{z_3}. \end{aligned}\right\} \tag{16}$$

Пример 2. С помощью способа Эйлера решить уравнение

$$x^4-6x^3+10x^2-2x-3=0.$$

Полагаем $x=y+\frac{3}{2}$, в результате чего получаем уравнение

$$y^4-\frac{7}{2}y^2+y+\frac{21}{16}=0. \tag{17}$$

Здесь $p=-\frac{7}{2}$, $q=1$ и $r=\frac{21}{16}$. Стало быть, уравнение (14)

принимает вид:

$$z^3 - 7z^2 + 7z - 1 = 0. \tag{18}$$

Корнями уравнения (18) являются

$$z_1 = 1, \quad z_2 = 3 + 2\sqrt{2}, \quad z_3 = 3 - 2\sqrt{2}$$

Так как $q = 1 > 0$, то для $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$ можно взять положительные значения. Таким образом,

$$2y_1 = 1 + \sqrt{3 + 2\sqrt{2}} + \sqrt{3 - 2\sqrt{2}},$$

$$2y_2 = 1 - \sqrt{3 + 2\sqrt{2}} - \sqrt{3 - 2\sqrt{2}},$$

$$2y_3 = -1 + \sqrt{3 + 2\sqrt{2}} - \sqrt{3 - 2\sqrt{2}},$$

$$2y_4 = -1 - \sqrt{3 + 2\sqrt{2}} + \sqrt{3 - 2\sqrt{2}},$$

или, так как $\sqrt{3 + 2\sqrt{2}} = 1 + \sqrt{2}$, $\sqrt{3 - 2\sqrt{2}} = 1 - \sqrt{2}$:

$$2y_1 = 3,$$
$$2y_2 = -1,$$
$$2y_3 = -1 + 2\sqrt{2},$$
$$2y_4 = -1 - 2\sqrt{2}.$$

Отсюда без труда находим корни x_1, x_2, x_3 и x_4 первоначального уравнения

$$x_1 = 3, \quad x_2 = 1, \quad x_3 = 1 + \sqrt{2}, \quad x_4 = 1 - \sqrt{2}.$$

§ 10. Алгебраическое расширение и другая постановка проблемы решения уравнений в радикалах

В тесной связи с проблемой решения уравнений в радикалах находится весьма важное понятие алгебраического расширения. К изложению этого понятия мы сейчас и перейдём.

Пусть P — некоторое числовое поле, α — произвольное комплексное число. Относительно α представляются, очевидно, только две возможности: либо α является корнем некоторого алгебраического уравнения n-й степени

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0$$

с коэффициентами из поля P, либо α не может быть корнем

никакого алгебраического уравнения произвольной степени с коэффициентами из поля P.

В первом случае α называется числом, *алгебраическим относительно поля* P, а во втором, в соответствии с общим определением трансцендентного элемента (см. § 1), — числом, трансцендентным относительно P. Отметим, что всякое a из поля P будет алгебраическим относительно P, так как оно является корнем уравнения $x-a=0$ с коэффициентами из поля P.

Если, в частности, P есть поле рациональных чисел, то слова «относительно поля P» обычно опускают и говорят просто об алгебраическом или трансцендентном числе. Например, $\sqrt{2}$ есть алгебраическое число (т. е. алгебраическое число относительно поля рациональных чисел), так как $\sqrt{2}$ является корнем уравнения второй степени $x^2-2=0$ с рациональными (а именно, целыми) коэффициентами.

Обозначим, далее, через $P[\alpha]$ множество чисел вида

$$f(\alpha)=c_0+c_1\alpha+c_2\alpha^2+\ldots+c_k\alpha^k,$$

где k — произвольное целое неотрицательное число, $c_0, c_1, \ldots, c_k$ — любые числа из поля P. Легко проверить, что это множество $P[\alpha]$ замкнуто относительно первых трёх арифметических действий и потому образует числовое кольцо.

Рассмотрим, кроме того, более обширное множество всевозможных отношений

$$\frac{f(\alpha)}{g(\alpha)}=\frac{c_0+c_1\alpha+\ldots+c_k\alpha^k}{d_0+d_1\alpha+\ldots+d_l\alpha^l} \qquad (g(\alpha)\neq 0)$$

элементов кольца $P[\alpha]$. Это множество, как легко видеть, является числовым полем. Мы его будем обозначать через $P(\alpha)$ (α в круглых скобках!) и называть *полем* или *простым расширением поля* P, *получающимся путём присоединения к* P *числа* α. При этом переход от поля P к полю $P(\alpha)$ называется *присоединением* элемента α к P.

Вообще, если некоторое числовое поле Δ содержится в некотором другом числовом поле Ω, то Δ, как известно, называется *подполем* поля Ω, а Ω — *расширением* поля Δ. Таким образом, простое расширение $P(\alpha)$ поля P есть частный случай понятия расширения поля.

Отметим ещё, что $P(\alpha)$ называется простым *алгебраическим* расширением поля P, если α является алгебраическим относительно P; если же α трансцендентно относительно P, то $P(\alpha)$ называется простым *трансцендентным* расширением P.

Пример 1. Возьмём в качестве поля P поле рациональных чисел и в качестве α — число $\sqrt{5}$. Выясним, что собой представляют $P[\sqrt{5}]$ и $P(\sqrt{5})$.

Согласно общему определению $P[\sqrt{5}]$ есть числовое кольцо, состоящее из элементов вида

$$c_0 + c_1(\sqrt{5}) + c_2(\sqrt{5})^2 + \ldots + c_k(\sqrt{5})^k, \tag{1}$$

где k — произвольное целое неотрицательное число, а $c_0, \ldots, c_k$ — произвольные рациональные числа.

Выражение (1) можно, однако, упростить. А именно, принимая во внимание, что

$$(\sqrt{5})^2 = 5, \quad (\sqrt{5})^3 = 5\sqrt{5}$$

и т. д., можно выражение (1) преобразовать в двучленное выражение $a + b\sqrt{5}$, где a, b — рациональные числа. Итак, мы видим, что $P[\sqrt{5}]$ состоит из всевозможных чисел вида $a + b\sqrt{5}$.

Теперь посмотрим, из чисел какого вида состоит $P(\sqrt{5})$. По определению $P(\sqrt{5})$ есть не что иное, как совокупность элементов вида

$$\frac{a + b\sqrt{5}}{c + d\sqrt{5}}, \tag{2}$$

где a, b, c, d — рациональные числа, $c + d\sqrt{5} \neq 0$. Для упрощения умножим числитель и знаменатель дроби (2) на $c - d\sqrt{5}$. Тогда получим:

$$\frac{a + b\sqrt{5}}{c + d\sqrt{5}} = a' + b'\sqrt{5},$$

где

$$a' = \frac{ac - 5bd}{c^2 - 5d^2}, \quad b' = \frac{bc - ad}{c^2 - 5d^2}.$$

Мы видим отсюда, что поле $P(\sqrt{5})$ состоит из элементов того же двучленного вида, что и кольцо $P[\sqrt{5}]$. Это значит, что кольцо $P[\sqrt{5}]$ совпадает с полем $P(\sqrt{5})$:

$$P[\sqrt{5}] = P(\sqrt{5}).$$

Только что подмеченное совпадение $P[\sqrt{5}]$ и $P(\sqrt{5})$ не является случайностью; оно характерно для всякого алгебраического числа.

Теорема 16. *Если α — число, алгебраическое относительно поля P, то уже $P[\alpha]$ является числовым полем*: $P[\alpha] = P(\alpha)$.

Доказательство. Без ограничения общности выводов можно предположить, что α является корнем многочлена

$$p(x)=p_0+p_1x+\ldots+p_nx^n \qquad (p_n\neq 0),$$

неприводимого в поле P. В самом деле, если бы α являлось корнем многочлена $F(x)$, приводимого в P, то, очевидно, α было бы корнем одного из неприводимых множителей $F(x)$. Этот неприводимый множитель мы и взяли бы в качестве $p(x)$.

Согласно определению любой элемент β из $P[\alpha]$ должен иметь вид

$$\beta=f(\alpha)=c_0+c_1\alpha+\ldots+c_k\alpha^k, \tag{3}$$

где k — целое неотрицательное число и $c_0, \ldots, c_k$ — числа из поля P. Покажем, прежде всего, что выражение (3) можно преобразовать в следующее выражение:

$$\beta=a_0+a_1\alpha+\ldots+a_{n-1}\alpha^{n-1},$$

имеющее относительно α степень, меньшую чем степень n многочлена $p(x)$.

Обозначим через $q(x)$ и $r(x)=a_0+a_1x+a_2x^2+\ldots+a_{n-1}x^{n-1}$ соответственно частное и остаток, получающиеся при делении

$$f(x)=c_0+c_1x+\ldots+c_kx^k$$

на $p(x)$. Мы можем написать, что

$$f(x)=p(x)q(x)+r(x). \tag{4}$$

Полагая в равенстве (4) $x=\alpha$ и принимая во внимание, что $p(\alpha)=0$, получаем: $f(\alpha)=r(\alpha)$, или

$$\beta=f(\alpha)=r(\alpha)=a_0+a_1\alpha+\ldots+a_{n-1}\alpha^{n-1},$$

что и требовалось показать.

Теперь обратимся к отношению

$$\frac{f(\alpha)}{g(\alpha)}=\frac{a_0+a_1\alpha+\ldots+a_{n-1}\alpha^{n-1}}{b_0+b_1\alpha+\ldots+b_{n-1}\alpha^{n-1}} \qquad (g(\alpha)\neq 0). \tag{5}$$

Мы сейчас убедимся, что его можно преобразовать в целое рациональное выражение от α. Для этой цели рассмотрим многочлен $g(x)=b_0+b_1x+\ldots+b_{n-1}x^{n-1}$. Он не равен (тождественно) нулю: если бы $g(x)$ равнялось нулю, то $b_0=b_1=\ldots=b_{n-1}=0$, в силу чего

$$g(\alpha)=b_0+b_1\alpha+\ldots+b_{n-1}\alpha^{n-1}=0,$$

что противоречит условию $g(\alpha)\neq 0$.

Очевидно, что $g(x)$ не делится на $p(x)$, так как степень $g(x)$ ниже степени $p(x)$. В силу неприводимости $p(x)$ отсюда следует, что многочлены $g(x)$ и $p(x)$ взаимно просты. Но в таком случае будет иметь место равенство

$$g(x)\varphi(x)+p(x)\psi(x)=1, \tag{6}$$

где $\varphi(x)$ и $\psi(x)$ — некоторые многочлены с коэффициентами из того же поля P (см. § 2). Полагая в равенстве (6) $x=\alpha$ и принимая во внимание, что $p(\alpha)=0$, получаем:

$$g(\alpha)\varphi(\alpha)=1. \tag{7}$$

Теперь, пользуясь равенством (7), мы можем следующим образом преобразовать отношение (5). Умножим числитель и знаменатель (5) на $\varphi(\alpha)$. Получим на основании равенства (7), что

$$\frac{f(\alpha)}{g(\alpha)}=\frac{f(\alpha)\varphi(\alpha)}{g(\alpha)\varphi(\alpha)}=\frac{f(\alpha)\varphi(\alpha)}{1}=f(\alpha)\varphi(\alpha).$$

Но $f(\alpha)\varphi(\alpha)$ есть некоторый многочлен от α:

$$f(\alpha)\varphi(\alpha)=c_0+c_1\alpha+\ldots+c_k\alpha^k.$$

Следовательно,

$$\frac{f(\alpha)}{g(\alpha)}=c_0+c_1\alpha+\ldots+c_k\alpha^k.$$

Итак, поле $P(\alpha)$ состоит из тех же элементов (3), что и кольцо $P[\alpha]$, вследствие чего $P(\alpha)=P[\alpha]$, и теорема доказана.

Замечание. В только что проведённом доказательстве мы попутно установили, что всякий элемент β из $P(\alpha)$ выражается в виде многочлена от α

$$\beta=a_0+a_1\alpha+\ldots+a_{n-1}\alpha^{n-1}$$

степени не выше, чем $n-1$, где n — степень многочлена $p(x)$. Нетрудно убедиться, что такое выражение элемента β является *единственным*. Действительно, если

$$\beta=b_0+b_1\alpha+\ldots+b_{n-1}\alpha^{n-1},$$

то

$$(a_0-b_0)+(a_1-b_1)\alpha+\ldots+(a_{n-1}-b_{n-1})\alpha^{n-1}=0.$$

Таким образом, многочлен $h(x)=(a_0-b_0)+(a_1-b_1)x+\ldots+(a_{n-1}-b_{n-1})x^{n-1}$ имеет α своим корнем, а потому многочлены $p(x)$ и $h(x)$ будут не взаимно просты, так как будут иметь $x-\alpha$ общим делителем. Тем самым многочлен $h(x)$ должен в силу неприводимости $p(x)$ делиться на $p(x)$. Но это может быть лишь тогда, когда $h(x)=0$; в противном случае $p(x)$ был бы делителем много-

члена $h(x)$ более низкой степени, что невозможно. Итак, $h(x)=0$, откуда $a_0-b_0=0, \ldots, a_{n-1}-b_{n-1}=0$, т. е. $a_0=b_0, \ldots, a_{n-1}=b_{n-1}$, и наше утверждение о единственности выражения β через α становится очевидным.

Подчёркиваем, что в случае трансцендентного α $P[\alpha]$ уже не будет полем.

В самом деле, если бы в этом случае $P[\alpha]$ являлось полем, то отношение $\frac{\alpha+1}{\alpha}$, в частности, было бы равно некоторому целому рациональному выражению от α:

$$\frac{\alpha+1}{\alpha}=c_0+c_1\alpha+\ldots+c_k\alpha^k,$$

где $c_0, \ldots, c_k$ — элементы из P. Отсюда мы имели бы, что

$$\alpha+1=c_0\alpha+c_1\alpha^2+\ldots+c_k\alpha^{k+1},$$

или

$$-1+(c_0-1)\alpha+c_1\alpha^2+\ldots+c_k\alpha^{k+1}=0,$$

что противоречит трансцендентности α, так как в последнем равенстве не все коэффициенты равны нулю (так, -1 отлична от нуля).

До сих пор мы присоединяли к полю P только одно число. Возьмём теперь несколько комплексных чисел $\alpha_1, \alpha_2, \ldots, \alpha_s$ (алгебраических или трансцендентных — безразлично). Присоединим сначала к P число α_1. Мы получим простое расширение $P(\alpha_1)$. Затем присоединим к $P(\alpha_1)$ число α_2. Получится дальнейшее расширение, которое мы обозначим через $P(\alpha_1, \alpha_2,)$, и т. д. После всех таких последовательных присоединений чисел $\alpha_1, \alpha_2, \ldots, \alpha_s$ мы придём к расширению $P(\alpha_1, \ldots, \alpha_s)$ поля P. Мы будем называть $P(\alpha_1, \ldots, \alpha_s)$ *расширением, полученным путём присоединения к P чисел* $\alpha_1, \ldots, \alpha_s$.

Оказывается, что числовое поле $P(\alpha_1, \ldots, \alpha_s)$ является минимальным среди всех числовых полей, содержащих P и $\alpha_1, \ldots, \alpha_s$, а именно, $P(\alpha_1, \ldots, \alpha_s)$ *есть пересечение всех числовых полей* Δ, *содержащих поле* P *и* $\alpha_1, \ldots, \alpha_s$.

Доказательство. Так как каждое из упомянутых полей Δ замкнуто относительно четырёх арифметических действий (исключая, конечно, деление на нуль), то Δ вместе с α_1 и P должно содержать и всевозможные числа вида

$$\frac{f(\alpha_1)}{g(\alpha_1)}=\frac{a_0+a_1\alpha_1+\ldots+a_k\alpha_1^k}{b_0+b_1\alpha_1+\ldots+b_l\alpha_1^l} \qquad (g(\alpha_1)\neq 0).$$

Это значит, что $P(\alpha_1)$ содержится в каждом из полей Δ. Но если $P(\alpha_1)$ содержится в Δ, то отсюда получается, что и $P(\alpha_1, \alpha_2)$ содержится в Δ. Действительно, поскольку Δ замкнуто относительно

четырёх арифметических действий, поле Δ должно содержать вместе с $P(\alpha_1)$ и α_2 и всевозможные числа вида

$$\frac{a_0(\alpha_1)+a_1(\alpha_1)\alpha_2+\ldots+a_k(\alpha_1)\alpha_2^k}{b_0(\alpha_1)+b_1(\alpha_1)\alpha_2+\ldots+b_l(\alpha_1)\alpha_2^l},$$

где $a_i(\alpha_1)$, $b_j(\alpha_1)$ — элементы из $P(\alpha_1)$ и знаменатель отличен от нуля. Иными словами, Δ должно содержать $P(\alpha_1, \alpha_2)$ и т. д. Рассуждая так, мы, наконец, убедимся, что Δ содержит $P(\alpha_1, \ldots, \alpha_s)$.

Обозначим теперь через Σ пересечение всех полей Δ, содержащих P и $\alpha_1, \ldots, \alpha_s$. Так как $P(\alpha_1, \ldots, \alpha_s)$ содержится во всех Δ, то

$$P(\alpha_1, \ldots, \alpha_s) \subseteq \Sigma. \tag{8}$$

С другой стороны, $P(\alpha_1, \ldots, \alpha_s)$, очевидно, содержит P и $\alpha_1, \ldots, \alpha_s$ и потому является одним из полей Δ, содержащих P и $\alpha_1, \ldots, \alpha_s$. Следовательно,

$$\Sigma \subseteq P(\alpha_1, \ldots, \alpha_s). \tag{9}$$

Сравнивая соотношения (8) и (9), мы видим, что $P(\alpha_1, \ldots, \alpha_s) = \Sigma$.

Из доказанной теоремы сразу вытекает, что расширение $P(\alpha_1, \ldots, \alpha_s)$ не зависит от того, в каком порядке присоединять к полю P числа $\alpha_1, \ldots, \alpha_s$:

$$P(\alpha_1, \ldots, \alpha_s) = P(\alpha_{i_1}, \ldots, \alpha_{i_s}).$$

В самом деле, $P(\alpha_1, \ldots, \alpha_s)$ есть пересечение всех числовых полей, содержащих P и $\alpha_1, \ldots, \alpha_s$; вследствие этого $P(\alpha_1, \ldots, \alpha_s)$ зависит лишь от поля P и от множества присоединяемых чисел $\alpha_1, \ldots, \alpha_s$, но не от порядка присоединения к полю P этих чисел.

Затем легко видеть, что $P(\alpha_1, \ldots, \alpha_s)$ есть не что иное, как совокупность чисел вида

$$\frac{A_1 \alpha_1^{k_1^{(1)}} \alpha_2^{k_2^{(1)}} \ldots \alpha_s^{k_s^{(1)}} + \ldots + A_p \alpha_1^{k_1^{(p)}} \alpha_2^{k_2^{(p)}} \ldots \alpha_s^{k_s^{(p)}}}{B_1 \alpha_1^{l_1^{(1)}} \alpha_2^{l_2^{(1)}} \ldots \alpha_s^{l_s^{(1)}} + \ldots + B_q \alpha_1^{l_1^{(p)}} \alpha_2^{l_2^{(p)}} \ldots \alpha_s^{l_s^{(p)}}}, \tag{10}$$

где A_i, B_j — элементы из поля P, а $k_i^{(\nu)}$ и $l_j^{(\mu)}$ — целые неотрицательные числа.

В самом деле, благодаря замкнутости $P(\alpha_1, \ldots, \alpha_s)$ относительно четырёх арифметических действий поле $P(\alpha_1, \ldots, \alpha_s)$ будет содержать вместе с P и $\alpha_1, \ldots, \alpha_s$ и всевозможные числа, получающиеся из чисел $\alpha_1, \ldots, \alpha_s$ и чисел из P с помощью той или иной комбинации четырёх арифметических действий. Короче говоря, $P(\alpha_1, \ldots, \alpha_s)$ должно содержать всевозможные числа вида (10). Но совокупность чисел вида (10) образует числовое поле. Следовательно, в силу минимальности, числами вида (10) должны исчерпываться все элементы поля $P(\alpha_1, \ldots, \alpha_s)$.

В дополнение к сказанному отметим, что если α_1 является алгебраическим относительно P, α_2—алгебраическим относительно $P(\alpha_1)$ и т. д., наконец, α_s является алгебраическим относительно $P(\alpha_1, \ldots, \alpha_{s-1})$, то $P(\alpha_1, \ldots, \alpha_s)$ будет исчерпываться числами вида

$$A_1 \alpha_1^{k_1^{(1)}} \alpha_2^{k_2^{(1)}} \ldots \alpha_s^{k_s^{(1)}} + \ldots + A_p \alpha_1^{k_1^{(p)}} \alpha_2^{k_2^{(p)}} \ldots \alpha_s^{k_s^{(p)}},$$

получающимися в результате комбинации первых трёх арифметических действий. Это следует из того, что в случае α алгебраического относительно поля P, кольцо $P[\alpha]$ является полем: $P[\alpha] = P(\alpha)$.

Проблема решения алгебраических уравнений в радикалах тесно связана с процессом расширения поля путём присоединения чисел, алгебраических относительно рассматриваемого поля. Этой связью мы воспользуемся при доказательстве теоремы Руффини-Абеля в § 16. Введём предварительно весьма важное понятие нормального поля.

Пусть $\alpha_1, \ldots, \alpha_n$ — все n корней уравнения

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n = 0 \qquad (11)$$

n-й степени. Присоединим к полю R рациональных чисел коэффициенты $a_0, a_1, \ldots, a_n$ уравнения. Мы получим расширение $R(a_0, a_1, \ldots, a_n)$, которое мы будем называть *областью рациональности уравнения* (11) и будем обозначать для краткости через Δ. Присоединим, далее, к Δ корни $\alpha_1, \alpha_2, \ldots, \alpha_n$. Полученное при этом расширение $\Delta(\alpha_1, \ldots, \alpha_n)$ поля Δ называется *нормальным полем* или *полем Галуа уравнения* (11). Нормальное поле $\Delta(\alpha_1, \ldots, \alpha_n)$ мы будем часто обозначать через Ω.

Теперь покажем, что уравнение (11) тогда и только тогда разрешимо в радикалах, когда нормальное поле $\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$ содержится в расширении $\Sigma = \Delta(\rho_1, \rho_2, \ldots, \rho_k)$, полученном путём присоединения к Δ некоторых радикалов $\rho_1 = \sqrt[n_1]{A_1}$, $\rho_2 = \sqrt[n_2]{A_2}$, $\ldots$ $\ldots, \rho_k = \sqrt[n_k]{A_k}$, где A_1 принадлежит Δ, A_2 принадлежит $\Delta(\rho_1)$, A_3 принадлежит $\Delta(\rho_1, \rho_2)$, $\ldots$, A_k принадлежит $\Delta(\rho_1, \ldots, \rho_{k-1})$.

В самом деле, если уравнение (11) разрешимо в радикалах, то это значит, что корни уравнения выражаются через его коэффициенты и некоторые радикалы $\rho_1, \rho_2, \ldots, \rho_k$ с помощью конечной комбинации четырёх арифметических действий. Так как поле $\Sigma = \Delta(\rho_1, \ldots, \rho_k)$ содержит коэффициенты $a_0, a_1, \ldots, a_n$ и радикалы $\rho_1, \rho_2, \ldots, \rho_k$ и, как и всякое числовое поле, замкнуто относительно арифметических действий, то корни $\alpha_1, \ldots, \alpha_n$ должны лежать в Σ. Следовательно, Ω, будучи минимальным среди всех числовых полей, содержащих Δ и $\alpha_1, \ldots, \alpha_n$, само должно содержаться в Σ.

Обратно, если Ω содержится в Σ: $\Omega \subseteq \Sigma$, то все корни $\alpha_1, \ldots, \alpha_n$ уравнения (11) лежат в Σ. Вследствие этого $\alpha_1, \ldots, \alpha_n$ будут выражаться через радикалы $\rho_1, \ldots, \rho_k$ и некоторые числа из Δ. Но каждое число из Δ в свою очередь выражается с помощью той или иной комбинации арифметических действий через коэффициенты уравнения (11), так как $\Delta = R(a_0, \ldots, a_n)$. Следовательно, корни уравнения (11) должны в конечном счёте выражаться через $a_0, \ldots, a_n$ и радикалы $\rho_1, \ldots, \rho_k$ с помощью некоторой конечной комбинации арифметических действий. Иными словами, уравнение (11) решается в радикалах.

Пример 2. Мы знаем, что уравнение третьей степени

$$x^3 + px + q = 0$$

можно решить по формулам

$$x_0 = u_0 - \frac{p}{3u_0}, \quad x_1 = u_1 - \frac{p}{3u_1}, \quad x_2 = u_2 - \frac{p}{3u_2}, \tag{12}$$

где u_0, u_1, u_2 суть значения

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Обозначим через A_1 выражение $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$. Очевидно, что A_1 лежит в $\Delta = R(p, q)$. Далее, обозначим через ρ_1 какое-нибудь из значений $\sqrt{A_1}$. Затем через A_2 обозначим:

$$-\frac{q}{2} + \rho_1.$$

Очевидно, что A_2 лежит в $\Delta(\rho_1)$. В качестве дальнейших радикалов мы возьмём $\rho_2 = u_0$, $\rho_3 = u_1$, $\rho_4 = u_2$. Таким образом, ρ_2, ρ_3, ρ_4 являются здесь значениями $\sqrt[3]{A_2}$. Теперь формулы (12) примут вид

$$x_0 = \rho_2 - \frac{p}{3\rho_2}, \quad x_1 = \rho_3 - \frac{p}{3\rho_3}, \quad x_2 = \rho_4 - \frac{p}{3\rho_4}. \tag{13}$$

Таким образом, корни x_0, x_1, x_2 выразились через радикалы ρ_1, ρ_2, ρ_3, ρ_4 и потому лежат в $\Delta(\rho_1, \rho_2, \rho_3, \rho_4)$.

Обращаем внимание читателя на следующее любопытное обстоятельство. Если уравнение (11) решается в радикалах, то всегда можно его корни выразить в таком виде, чтобы радикалы $\rho_1, \rho_2, \ldots, \rho_k$ не входили в знаменатели, т. е. чтобы корни получались из $\rho_1, \rho_2, \ldots, \rho_k$ и некоторых чисел поля Δ с помощью действий сложения, вычитания и умножения, но не деления. Это следует из того, что ρ_1 является алгебраическим относительно Δ, ρ_2 является

алгебраическим относительно $\Delta(\rho_1)$ и т. д. Например, в формулах (13) только что разобранного примера радикалы ρ_2, ρ_3, ρ_4 встречаются и в знаменателях; однако мы можем формулы (13) преобразовать так, чтобы знаменатели зависели только от элементов поля Δ. А именно, если под ρ понимать любой из радикалов ρ_2, ρ_3, ρ_4, то

$$\frac{1}{\rho}=\frac{\rho^2}{\rho^3}=\frac{\rho^2}{-\frac{q}{2}+\rho_1}=\frac{\rho^2\left(-\frac{q}{2}-\rho_1\right)}{\left(\frac{q}{2}\right)^2-\rho_1^2}=\frac{\rho^2\left(-\frac{q}{2}-\rho_1\right)}{\left(\frac{q}{2}\right)^2-\left(\frac{q}{2}\right)^2-\left(\frac{p}{3}\right)^3}=$$

$$=\frac{27\rho^2\left(\frac{q}{2}+\rho_1\right)}{p^3},$$

откуда

$$\left.\begin{aligned} x_0&=\rho_2-\frac{9\rho_2^2\left(\frac{q}{2}+\rho_1\right)}{p^2},\\ x_1&=\rho_3-\frac{9\rho_3^2\left(\frac{q}{2}+\rho_1\right)}{p^2},\\ x_2&=\rho_4-\frac{9\rho_4^2\left(\frac{q}{2}+\rho_1\right)}{p^2}. \end{aligned}\right\}$$

ГЛАВА II

КОЛЬЦО МНОГОЧЛЕНОВ ОТ НЕСКОЛЬКИХ НЕИЗВЕСТНЫХ И ПОЛЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ

§ 11. Кольцо многочленов от нескольких неизвестных

Пользуясь понятием кольца многочленов от одного неизвестного, мы сейчас определим индуктивно кольцо многочленов от нескольких неизвестных. А именно, пусть R — снова коммутативное кольцо с единицей $e \neq 0$. Тогда кольцо многочленов от x_2 над R_1, где R_1 в свою очередь есть кольцо многочленов от x_1 над R, мы назовём кольцом многочленов от двух неизвестных x_1 и x_2 над R и обозначим через $R[x_1, x_2]$.

Вообще кольцо многочленов от x_n над R_{n-1}, где R_{n-1} — кольцо многочленов от $x_1, \ldots, x_{n-1}$ над R, мы будем называть кольцом многочленов от n неизвестных $x_1, x_2, \ldots, x_n$ и будем его обозначать через $R[x_1, x_2, \ldots, x_n]$.

Посмотрим теперь, что представляют собою элементы кольца $R[x_1, \ldots, x_n]$. Обратимся сначала к кольцу $R[x_1, x_2]$ от двух неизвестных. Каждый элемент f этого кольца является не чем иным, как многочленом от x_2 над $R[x_1]$:

$$f = a_0(x_1) + a_1(x_1)\,x_2 + \ldots + a_m(x_1)\,x_2^m, \tag{1}$$

причём каждый из коэффициентов $a_i(x_1)$ является многочленом от x_1 над R:

$$\left.\begin{aligned} a_i(x_1) &= a_{i0} + a_{i1}x_1 + \ldots + a_{im_i}x_1^{m_i} \\ (i &= 0,\ 1,\ \ldots,\ m;\ a_{ik} \text{ — элементы из } R). \end{aligned}\right\} \tag{2}$$

Сложение и умножение, употребляемые в выражениях (1) и (2), следует понимать в смысле операций сложения и умножения кольца $R[x_1, x_2]$, так как $R[x_1]$ есть подкольцо кольца $R[x_1, x_2]$. Благодаря этому мы вправе заменить в равенстве (1) коэффициенты $a_i(x_1)$ их выражениями (2) и раскрыть скобки. В результате таких преобразований элемент f выразится в виде конечной суммы

$$f = A_1x_1^{\alpha_1}x_2^{\beta_1} + A_2x_1^{\alpha_2}x_2^{\beta_2} + \ldots + A_kx_1^{\alpha_k}x_2^{\beta_k} \qquad (k \geqslant 1), \tag{3}$$

где α_i, β_i — целые неотрицательные числа, A_j — элементы из R. Можно всегда предполагать, что в сумме (3) нет подобных слагаемых, т. е. слагаемых, отличающихся друг от друга лишь множителем A_j. В самом деле, если бы, например, $A_1 x_1^{\alpha_1} x_2^{\beta_1}$ и $A_2 x_1^{\alpha_2} x_2^{\beta_2}$ были подобны, то $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, вследствие чего

$$A_1 x_1^{\alpha_1} x_2^{\beta_1} + A_2 x_1^{\alpha_2} x_2^{\beta_2} = (A_1 + A_2) x_1^{\alpha_1} x_2^{\beta_1},$$

т. е. мы могли бы $A_1 x_1^{\alpha_1} x_2^{\beta_1}$ и $A_2 x_1^{\alpha_2} x_2^{\beta_2}$ заменить одним слагаемым $(A_1 + A_2) x_1^{\alpha_1} x_2^{\beta_1}$.

Итак, каждый элемент f кольца $R[x_1, x_2]$ представляется в виде выражения (3). Покажем, что и в общем случае n ($n \geqslant 1$) неизвестных наблюдается та же картина: *каждый элемент f кольца $R[x_1, \ldots, x_n]$ можно представить в виде выражения*

$$f = A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k}, \tag{4}$$

причём $\alpha_i, \beta_i, \ldots, \omega_i$ — целые неотрицательные числа, $A_1, A_2, \ldots, A_k$ — элементы из R, и в выражении (4) *нет подобных слагаемых.*

Для этой цели воспользуемся методом математической индукции: допустим, что наше утверждение справедливо для $n-1$ неизвестных, и покажем, что тогда оно будет верно и для n неизвестных.

Действительно, так как f есть многочлен от x_n над кольцом $R[x_1, \ldots, x_{n-1}]$, то

$$f = a_0(x_1, \ldots, x_{n-1}) + a_1(x_1, \ldots, x_{n-1}) x_n + \ldots + a_m(x_1, \ldots, x_{n-1}) x_n^m, \tag{5}$$

где $a_i(x_1, \ldots, x_{n-1})$ — многочлен от $n-1$ неизвестных $x_1, \ldots, x_{n-1}$ над R. В силу нашего допущения все коэффициенты $a_i(x_1, \ldots, x_{n-1})$ можно представить в виде выражений

$$a_i(x_1, \ldots, x_{n-1}) = A_{i1} x_1^{\mu_1} x_2^{\nu_1} \ldots x_n^{\rho_1} + \ldots + A_{is} x_1^{\mu_s} x_2^{\nu_s} \ldots x_n^{\rho_s}. \tag{6}$$

Но $R[x_1, \ldots, x_{n-1}]$ есть подкольцо кольца $R[x_1, \ldots, x_n]$. Следовательно, сложение и умножение, употребляемые в выражениях (5) и (6), надо понимать в смысле операций сложения и умножения кольца $R[x_1, \ldots, x_n]$. Таким образом, в правой части равенства (5) можно коэффициенты $a_i(x_1, \ldots, x_{n-1})$ заменить их выражениями (6), раскрыть скобки и произвести приведение подобных членов; в результате у нас как раз и получится выражение (4) элемента f. Тем самым наше утверждение доказано для любого n, так как оно верно для случая одного неизвестного.

Очевидно, что справедливо и обратное: *всякая сумма вида* (4) *представляет собою элемент f кольца $R[x_1, \ldots, x_n]$.*

В самом деле, каждое слагаемое $A_i x_1^{\alpha_i} x_2^{\beta_i} \ldots x_n^{\omega_i}$ есть многочлен от x_n над $R[x_1, \ldots, x_{n-1}]$, так как $A_i x_1^{\alpha_i} \ldots x_{n-1}^{\varkappa_i}$ лежит в $R[x_1, \ldots, x_{n-1}]$. Поэтому сумма таких слагаемых будет также много-

членом от x_n над $R[x_1, \ldots, x_{n-1}]$, т. е. будет элементом кольца $R[x_1, \ldots, x_n]$.

Элементы f кольца $R[x_1, \ldots, x_n]$ называются *многочленами от неизвестных* $x_1, \ldots, x_n$ над R и обозначаются через $f(x_1, \ldots, x_n)$, $g(x_1, \ldots, x_n)$ и т. п. Само выражение (4) обычно называется *нормальным видом* многочлена, слагаемое $A_i x_1^{\alpha_i} \ldots x_n^{\omega_i}$ — *членом* и A_i — *коэффициентом* многочлена.

В частности, всякий элемент a из кольца R можно рассматривать как многочлен от n неизвестных над R, причём в случае $a \neq 0$ следует предполагать, что показатели неизвестных равны нулю.

Пользуясь общими свойствами операций сложения, вычитания и умножения элементов кольца, нетрудно получить обычные правила обращения с многочленами от нескольких неизвестных. А именно, пусть

$$f(x_1, \ldots, x_n) = A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k},$$

$$g(x_1, \ldots, x_n) = B_1 x_1^{\alpha'_1} x_2^{\beta'_1} \ldots x_n^{\omega'_1} + \ldots + B_l x_1^{\alpha'_l} x_2^{\beta'_l} \ldots x_n^{\omega'_l}$$

— два каких-нибудь многочлена из $R[x_1, \ldots, x_n]$. Тогда

$$f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n) = (A_1 x_1^{\alpha_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} \ldots x_n^{\omega_k}) + (B_1 x_1^{\alpha'_1} \ldots x_n^{\omega'_1} + \ldots + B_l x_1^{\alpha'_l} \ldots x_n^{\omega'_l}).$$

Так как во всяком кольце и, в частности в кольце $R[x_1, \ldots, x_n]$, сложение подчиняется сочетательному и переместительному законам и имеет место распределительный закон, то мы можем раскрыть скобки и произвести приведение подобных членов, в результате чего получится нормальный вид суммы $f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n)$.

Что касается разности $f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$, то она сводится к сумме $f(x_1, \ldots, x_n) + [-g(x_1, \ldots, x_n)]$, причём

$$-g(x_1, \ldots, x_n) = (-B_1) x_1^{\alpha'_1} x_2^{\beta'_1} \ldots x_n^{\omega'_1} + \ldots + (-B_l) x_1^{\alpha'_l} x_2^{\beta'_l} \ldots x_n^{\omega'_l}.$$

Для получения произведения $f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n)$ можно воспользоваться правилом умножения суммы на сумму

$$(a_1 + a_2 + \ldots + a_p)(b_1 + b_2 + \ldots + b_q) = a_1 b_1 + \ldots + a_1 b_q + \ldots + a_p b_q,$$

имеющим место во всяком кольце. Согласно этому правилу надо каждый член

$$A_i x_1^{\alpha_i} x_2^{\beta_i} \ldots x_n^{\omega_i}$$

многочлена $f(x_1, \ldots, x_n)$ умножить на каждый член

$$B_j x_1^{\alpha'_j} x_2^{\beta'_j} \ldots x_n^{\omega'_j}$$

многочлена $g(x_1, \ldots, x_n)$; затем, пользуясь сочетательностью и переместительностью умножения, а также распределительным законом, можно получившиеся произведения

$$(A_i x_1^{\alpha_i} x_2^{\beta_i} \ldots x_n^{\omega_i})(B_j x_1^{\alpha'_j} x_2^{\beta'_j} \ldots x_n^{\omega'_j})$$

преобразовать в

$$A_i B_j x_1^{\alpha_i + \alpha'_j} x_2^{\beta_i + \beta'_j} \ldots x_n^{\omega_i + \omega'_j}$$

и затем произвести приведение подобных членов. В результате получится нормальный вид произведения $f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n)$.

Мы знаем, что многочлен над R от одного неизвестного равен нулю лишь в том случае, когда равны нулю все его коэффициенты. Оказывается, что это обстоятельство справедливо и для многочленов от n неизвестных, т. е. имеет место

Теорема 17. *Многочлен*

$$f(x_1, \ldots, x_n) = A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k} \qquad (7)$$

из $R[x_1, \ldots, x_n]$ тогда и только тогда равен нулю, когда все его коэффициенты равны нулю.

Доказательство. Если все коэффициенты A_i многочлена (7) равны нулю, то, очевидно, многочлен сам равен нулю.

Обратно, пусть

$$A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k} = 0. \qquad (8)$$

Поскольку теорема верна для многочлена от одного неизвестного над R, воспользуемся методом математической индукции: допустим, что теорема верна для многочленов от $n-1$ неизвестных $x_1, \ldots, x_{n-1}$ над R, и покажем, что тогда теорема будет верна и для многочлена от n неизвестных.

Выносим за скобки каждую степень неизвестного x_n в левой части равенства (8). Получаем:

$$a_1(x_1, \ldots, x_{n-1}) x_n^{m_1} + \ldots + a_s(x_1, \ldots, x_{n-1}) x_n^{m_s} = 0, \qquad (9)$$

где $a_i(x_1, \ldots, x_{n-1})$ — многочлены уже от $n-1$ неизвестных $x_1, \ldots, x_{n-1}$ над R. Рассматривая левую часть равенства (9) как многочлен от одного неизвестного x_n над кольцом $R(x_1, \ldots, x_{n-1})$ и пользуясь тем, что теорема верна для случая одного неизвестного, получаем:

$$a_1(x_1, \ldots, x_{n-1}) = \ldots = a_s(x_1, \ldots, x_{n-1}) = 0.$$

Но мы допустили, что теорема верна и для случая $n-1$ неизвестных. Следовательно, коэффициенты многочленов

$$a_1(x_1, \ldots, x_{n-1}), \quad \ldots, \quad a_s(x_1, \ldots, x_{n-1})$$

все должны равняться нулю и тем самым должны равняться нулю коэффициенты многочлена $f(x_1, \ldots, x_n)$. Теорема доказана.

Таким образом, если

$$f(x_1, \ldots, x_n) = A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k} \neq 0,$$

то по меньшей мере один из коэффициентов $A_1, \ldots, A_k$ должен быть отличен от нуля. В дальнейшем мы будем в нормальном виде многочлена, отличного от нуля, опускать члены с нулевыми коэффициентами.

Из только что доказанной теоремы 17 вытекает единственность представления многочлена в нормальном виде, а именно вытекает

Теорема 18. *Два многочлена $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ из $R[x_1, \ldots, x_n]$ равны тогда и только тогда, когда члены многочлена $f(x_1, \ldots, x_n)$ входят в состав членов многочлена $g(x_1, \ldots, x_n)$, и обратно, члены многочлена $g(x_1, \ldots, x_n)$ входят в состав членов многочлена $f(x_1, \ldots, x_n)$.*

Доказательство. Если два многочлена $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ по составу своих членов не отличаются друг от друга, то

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n),$$

так как многочлены представляют собой сумму одних и тех же элементов кольца.

Обратно, пусть

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n).$$

Предположим, например, что многочлен $f(x_1, \ldots, x_n)$ имеет член, не входящий в состав членов многочлена $g(x_1, \ldots, x_n)$. Тогда разность $f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$ будет содержать по меньшей мере один член с коэффициентом, отличным от нуля, и в то же время будет равна нулю:

$$f - g = C_1 x_1^{\mu_1} x_2^{\nu_1} \ldots x_n^{\rho_1} + \ldots + C_s x_1^{\mu_s} x_2^{\nu_s} \ldots x_n^{\rho_s} = 0 \quad (C_1 \neq 0). \quad (10)$$

Но равенство (10) противоречит теореме 17.

Для изучения дальнейших свойств многочлена от нескольких неизвестных введём понятие степени многочлена.

Степенью многочлена $f(x_1, \ldots, x_n) \neq 0$ из кольца $R[x_1, \ldots, x_n]$ по отношению к неизвестному x_i называется наибольший показатель, с которым x_i входит в члены многочлена. Например, степень многочлена

$$x_1^2 x_2^3 + 5x_1 x_3 - x_2 x_3^6 + 1$$

над кольцом целых чисел относительно x_1 равна двум, относительно x_2 равна трём и относительно x_3 равна шести.

Если в многочлен $f(x_1, \ldots, x_n) \neq 0$ неизвестное x_i фактически не входит, то степень многочлена $f(x_1, \ldots, x_n)$ относительно этого неизвестного x_i будет, очевидно, равна нулю.

Назовём, далее, *степенью члена*

$$A_i x_1^{\alpha_i} x_2^{\beta_i} \ldots x_n^{\omega_i}$$

многочлена $f(x_1, \ldots, x_n) \neq 0$ сумму показателей $\alpha_i + \beta_i + \ldots + \omega_i$ неизвестных. Тогда *степенью многочлена* (по отношению ко всей совокупности неизвестных) называется наибольшая из степеней его членов.

Так, например, степень многочлена

$$2x_1^5 + x_1 x_2^4 + x_1 x_3^9 + 1$$

над кольцом целых чисел равна десяти.

Мы ввели понятие степени для многочленов $f(x_1, \ldots, x_n)$ из $R(x_1, \ldots, x_n]$, отличных от нуля. Что касается многочлена, равного нулю, то мы его будем считать многочленом, не имеющим степени. Очевидно, что всякий элемент $a \neq 0$ кольца R можно рассматривать как многочлен нулевой степени над R от неизвестных $x_1, \ldots, x_n$. В этом отношении мы имеем здесь ту же картину, что и в случае многочлена от одного неизвестного. Однако в отличие от случая многочлена от одного неизвестного здесь уже нельзя говорить в прежнем понимании о старшем члене, так как в многочлене от многих неизвестных могут встретиться несколько членов наибольшей степени, а в некоторых случаях все члены многочлена могут быть одной и той же степени.

Например, степень многочлена

$$x_1^5 + x_1 x_2^4 - 3x_1^2 x_2 + 5x_2^2 - 1$$

равна пяти, и многочлен имеет два члена x_1^5 и $x_1 x_2^4$ со степенью, равной пяти. В многочлене

$$5xyz - x^2 y + y^2 z \tag{11}$$

все члены имеют третью степень.

Многочлен $f(x_1, \ldots, x_n) \neq 0$ из $R[x_1, \ldots, x_n]$ мы назовём *однородным* или *формою* m-й степени, если все его члены имеют одну и ту же степень, равную m. В частности, форма первой степени часто называется линейной, второй степени — квадратичной, третьей степени — кубичной. Рассмотренный выше многочлен (11) является кубичной формой от трёх неизвестных.

Если кольцо R не содержит делителей нуля, то справедлива следующая

Теорема 19. *Если кольцо R не содержит делителей нуля, то кольцо $R[x_1, \ldots, x_n]$ также не содержит делителей нуля.*

Доказательство. Мы знаем, что теорема верна для случая одного неизвестного (см. § 1). Поэтому воспользуемся методом ма-

тематической индукции; допустим, что в кольце многочленов $R[x_1, \ldots, x_{n-1}]$ от $n-1$ неизвестных нет делителей нуля. Тогда и в кольце $R[x_1, \ldots, x_n]$ также не будет делителей нуля, так как по определению $R[x_1, \ldots, x_n]$ есть кольцо многочленов от одного неизвестного x_n над кольцом $R[x_1, \ldots, x_{n-1}]$ без делителей нуля.

Исходя из только что доказанной теоремы, нетрудно убедиться, что теорема о степени произведения двух многочленов от одного неизвестного, высказанная нами ещё в § 1, может быть распространена и на многочлены от нескольких неизвестных:

Теорема 20. *Если кольцо R не содержит делителей нуля, то степень произведения двух многочленов из $R[x_1, \ldots, x_n]$ равна сумме их степеней.*

Доказательство. Покажем, прежде всего, что теорема верна для однородных многочленов.

Пусть $\varphi(x_1, \ldots, x_n)$ и $\psi(x_1, \ldots, x_n)$ — две какие-нибудь формы из $R[x_1, \ldots, x_n]$ соответственно m_1-й и m_2-й степени:

$$\varphi = A_1 x_1^{\alpha_1} x_2^{\beta_1} \ldots x_n^{\omega_1} + \ldots + A_k x_1^{\alpha_k} x_2^{\beta_k} \ldots x_n^{\omega_k},$$

$$\psi = B_1 x_1^{\alpha'_1} x_2^{\beta'_1} \ldots x_n^{\omega'_1} + \ldots + B_l x_1^{\alpha'_l} x_2^{\beta'_l} \ldots x_n^{\omega'_l},$$

$$\alpha_i + \beta_i + \ldots + \omega_i = m_1; \quad \alpha'_j + \beta'_j + \ldots + \omega'_j = m_2$$

$$(i = 1, 2, \ldots, k; \; j = 1, 2, \ldots, l).$$

Умножая каждый член формы φ на каждый член формы ψ, мы получим сумму слагаемых вида

$$A_i B_j x_1^{\alpha_i + \alpha'_j} x_2^{\beta_i + \beta'_j} \ldots x_n^{\omega_i + \omega'_j}, \tag{12}$$

и степень каждого из таких слагаемых по отношению ко всей совокупности неизвестных равна

$$(\alpha_i + \alpha'_j) + (\beta_i + \beta'_j) + \ldots + (\omega_i + \omega'_j) =$$
$$= (\alpha_i + \beta_i + \ldots + \omega_i) + (\alpha'_j + \beta'_j + \ldots + \omega'_j) = m_1 + m_2.$$

Следовательно, если $\varphi\psi \neq 0$, то не все слагаемые (12) исчезнут, и степень $\varphi\psi$ будет равна $m_1 + m_2$. Но $\varphi\psi$ не может равняться нулю, так как $R[x_1, \ldots, x_n]$ есть кольцо без делителей нуля в силу теоремы 19. Итак, для форм наша теорема доказана.

Остаётся доказать теорему для произвольных многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ из $R[x_1, \ldots, x_n]$.

Пусть степень многочлена $f(x_1, \ldots, x_n)$ равна m_1, а степень многочлена $g(x_1, \ldots, x_n)$ равна m_2. Тогда мы можем написать,

что

$$f(x_1, \ldots, x_n) = \varphi_{m_1}(x_1, \ldots, x_n) + \varphi_{k_1}(x_1, \ldots, x_n) + \ldots + \varphi_{r_1}(x_1, \ldots, x_n),$$

$$g(x_1, \ldots, x_n) = \psi_{m_2}(x_1, \ldots, x_n) + \psi_{k_2}(x_1, \ldots, x_n) + \ldots + \psi_{r_2}(x_1, \ldots, x_n),$$

где φ_{m_1}, φ_{k_1}, ..., φ_{r_1} — формы, соответственно, m_1-й, k_1-й, ..., r_1-й степени, ψ_{m_2}, ψ_{k_2}, ..., ψ_{r_2} — формы, соответственно, m_2-й, k_2-й, ..., r_2-й степени и $m_1 > k_1 > \ldots > r_1$; $m_2 > k_2 > \ldots > r_2$. Перемножая $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$, получаем:

$$f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n) = \varphi_{m_1}(x_1, \ldots, x_n)\, \psi_{m_2}(x_1, \ldots, x_n) + \ldots + \varphi_{r_1}(x_1, \ldots, x_n)\, \psi_{r_2}(x_1, \ldots, x_n).$$

Очевидно, что в произведении $f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n)$ наибольшую степень имеют члены, входящие в

$$\varphi_{m_1}(x_1, \ldots, x_n)\, \psi_{m_2}(x_1, \ldots, x_n). \tag{13}$$

Но по доказанному выше степень произведения (13) форм $\varphi_{m_1}(x_1, \ldots, x_n)$ и $\psi_{m_2}(x_1, \ldots, x_n)$ должна равняться сумме $m_1 + m_2$ степеней этих форм. Следовательно, степень произведения многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ также равна $m_1 + m_2$.

Только что доказанную теорему можно распространить и на произведение нескольких многочленов: *степень произведения S многочленов из $R[x_1, \ldots, x_n]$ равна сумме степеней этих многочленов.*

Понятие значения многочлена от нескольких неизвестных вводится совершенно так же, как и в случае многочлена от одного неизвестного. Именно, пусть $f(x_1, \ldots, x_n)$ — произвольный многочлен из $R[x_1, \ldots, x_n]$. Заменим в нём неизвестные $x_1, \ldots, x_n$ какими-нибудь элементами $c_1, \ldots, c_n$ кольца R. Мы получим некоторый элемент d того же кольца R. Этот элемент и называется значением многочлена $f(x_1, \ldots, x_n)$ при значениях неизвестных $x_1 = c_1, \ldots, x_n = c_n$ и обозначается через $f(c_1, \ldots, c_n)$.

Очевидно, что если два многочлена из $R[x_1, \ldots, x_n]$ равны, то их значения также равны при любых значениях неизвестных. Для произвольного кольца R обратное, как мы знаем, неверно уже в случае многочлена от одного неизвестного. Однако дело обстоит иначе, когда кольцо R бесконечно и не обладает делителями нуля. Докажем, прежде всего, следующую теорему:

Теорема 21. *Если кольцо R бесконечно и не содержит делителей нуля, то многочлен $f(x_1, \ldots, x_n)$ над R в том и только в том случае равен нулю, когда при любых значениях неизвестных он обращается в нуль.*

Доказательство. Если многочлен $f(x_1, \ldots, x_n)$ равен нулю, то все его коэффициенты должны равняться нулю, и

потому многочлен будет обращаться в нуль при любых значениях неизвестных.

Обратно, пусть многочлен $f(x_1, \ldots, x_n)$ обращается в нуль при любых значениях неизвестных. Для случая одного неизвестного (т. е. при $n=1$) теорема была доказана уже в § 3. Поэтому воспользуемся методом математической индукции: предположим, что теорема верна для случая $n-1$ неизвестных, и покажем, что тогда теорема будет верна и для случая n неизвестных.

Многочлен $f(x_1, \ldots, x_n)$ можно записать в виде

$$f(x_1, \ldots, x_n) = a_0(x_1, \ldots, x_{n-1}) + a_1(x_1, \ldots, x_{n-1})x_n + \\ + \ldots + a_m(x_1, \ldots, x_{n-1})x_n^m,$$

где $a_i(x_1, \ldots, x_n)$ — многочлены от $n-1$ неизвестных $x_1, \ldots, x_{n-1}$. Дадим неизвестным $x_1, x_2, \ldots, x_{n-1}$ произвольные значения $b_1, b_2, \ldots, b_{n-1}$. Тогда мы получим многочлен уже от одного неизвестного x_n над кольцом R:

$$a_0(b_1, \ldots, b_{n-1}) + a_1(b_1, \ldots, b_{n-1})x_n + \ldots + a_m(b_1, \ldots, b_{n-1})x_n^m. \quad (14)$$

Так как $f(x_1, \ldots, x_n)$ равен нулю при любых значениях неизвестных, то многочлен (14) будет равен нулю при любом значении неизвестного x_n. Отсюда в силу того, что теорема верна для многочлена от одного неизвестного над R, получаем:

$$a_0(b_1, \ldots, b_{n-1}) = 0, \quad \ldots, \quad a_m(b_1, \ldots, b_{n-1}) = 0. \quad (15)$$

Равенства (15) свидетельствуют о том, что многочлены $a_i(x_1, \ldots, x_{n-1})$ от $n-1$ неизвестных $x_1, \ldots, x_{n-1}$ над R обращаются в нуль при любых значениях неизвестных, так как $b_1, \ldots, b_{n-1}$ произвольны. Но, по предположению, теорема верна для многочленов от $n-1$ неизвестных. Следовательно,

$$a_0(x_1, \ldots, x_{n-1}) = 0, \quad \ldots, \quad a_m(x_1, \ldots, x_{n-1}) = 0,$$

в силу чего и $f(x_1, \ldots, x_n)$ равен нулю.

Из только что доказанной теоремы вытекает

Теорема 22. *Если кольцо R бесконечно и не содержит делителей нуля, то два многочлена из $R[x_1, \ldots, x_n]$ равны в том и только в том случае, когда их значения совпадают при любых значениях неизвестных.*

Доказательство. Если многочлены $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ равны, то, как мы уже выше отметили, их значения совпадают независимо от того, будет или не будет R бесконечным кольцом без делителей нуля. Поэтому рассмотрим обратное: пусть значения многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ совпадают при любых значениях неизвестных. Тогда разность

$$f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$$

будет обращаться в нуль при любых значениях неизвестных и потому на основании теоремы 21 будет равна нулю:

$$f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n) = 0,$$

откуда $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)$.

В заключение этого параграфа отметим следующее: с помощью примерно тех же соображений, что и для случая многочлена от одного неизвестного, нетрудно убедиться, что если

$$f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n) = h(x_1, \ldots, x_n),$$
$$f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n) = k(x_1, \ldots, x_n),$$

то

$$\left.\begin{aligned} f(c_1, \ldots, c_n) + g(c_1, \ldots, c_n) &= h(c_1, \ldots, c_n), \\ f(c_1, \ldots, c_n)\, g(c_1, \ldots, c_n) &= k(c_1, \ldots, c_n) \end{aligned}\right\} \qquad (16)$$

при произвольных значениях $x_1 = c_1, \ldots, x_n = c_n$ неизвестных.

Пользуясь соотношениями (16) и теоремой 22, можно затем убедиться, что для бесконечного кольца R без делителей нуля алгебраическая и функциональная точки зрения на многочлен от нескольких неизвестных равносильны. Рассуждения по существу будут теми же, что и для многочлена от одного неизвестного.

§ 12. Поле алгебраических дробей

Многочлены являются частным случаем понятия алгебраической дроби. В этом параграфе мы сначала дадим соответствующее определение алгебраической дроби, а затем выясним, при каких условиях можно эти дроби рассматривать как функции.

Пусть P — произвольное поле. Множество многочленов $P[x_1, \ldots, x_n]$ от неизвестных $x_1, \ldots, x_n$ образует, как мы знаем из предыдущего параграфа, коммутативное кольцо без делителей нуля. Однако $P[x_1, \ldots, x_n]$ полем всё же не является, так как далеко не всегда один многочлен делится на другой. Например, $x_1^2 + a$, где $a \neq 0$ — элемент из P, не делится на $x_1^3 + a$; не делится по той простой причине, что степень делимого меньше степени делителя.

Предположим теперь, что существует такое поле Ω, в котором кольцо $P[x_1, \ldots, x_n]$ является подкольцом. Тогда для каждой пары многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n) \neq 0$ из $P[x_1, \ldots, x_n]$ уравнение

$$g(x_1, \ldots, x_n)\, z = f(x_1, \ldots, x_n)$$

будет иметь в поле Ω единственное решение $z = \alpha$. Обозначим это решение через

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}$$

и назовём *алгебраической дробью* от неизвестных $x_1, \ldots, x_n$ над полем P. При этом $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$ будут называться соответственно числителем и знаменателем дроби.

На основании общих свойств поля легко установить, что действия над алгебраическими дробями ничем не отличаются от действий над обычными дробями, а именно:

1°.

$$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} = \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)}$$

тогда и только тогда, когда

$$f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n) = f_2(x_1, \ldots, x_n) g_1(x_1, \ldots, x_n).$$

2°.

$$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} + \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n) + f_2(x_1, \ldots, x_n) g_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)}$$

3°.

$$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} \cdot \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) f_2(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)}.$$

4°.

$$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} : \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) f_2(x_1, \ldots, x_n)}$$

$$(f_2(x_1, \ldots, x_n) \neq 0).$$

Мы ограничимся выводом свойства 1°, причём для сокращения письма будем многочлены $f(x_1, \ldots, x_n)$, $g(x_1, \ldots, x_n), \ldots$ обозначать одной буквой: $f, g, \ldots$

Если

$$\frac{f_1}{g_1} = \frac{f_2}{g_2} = \alpha,$$

то по определению алгебраической дроби

$$g_1\alpha = f_1, \quad g_2\alpha = f_2.$$

Умножим первое равенство на g_2, а второе на g_1. Получим:

$$g_1 g_2 \alpha = f_1 g_2, \quad g_1 g_2 \alpha = f_2 g_1,$$

откуда $f_1 g_2 = f_2 g_1$.

Обратно, пусть $f_1 g_2 = f_2 g_1$ и $g_1\alpha = f_1$. Покажем, что это α будет также решением и уравнения $g_2 z = f_2$.

Умножаем обе части равенства $g_1\alpha = f_1$ на g_2. Получаем:

$$g_1 g_2 \alpha = f_1 g_2.$$

Но $f_1 g_2 = f_2 g_1$. Следовательно, заменяя в правой части равенства $f_1 g_2$ выражением $f_2 g_1$, имеем:

$$g_1 g_2 \alpha = f_2 g_1.$$

Последнее равенство можно сократить на g_1, так как $g_1 \neq 0$. Получаем в результате такого сокращения, что $g_2\alpha = f_2$. Отсюда вытекает, что $\alpha = \frac{f_2}{g_2}$, т. е. $\frac{f_2}{g_2} = \frac{f_1}{g_1}$.

Из свойств 1°—4° вытекает, что множество всех алгебраических дробей над P от неизвестных $x_1; \ldots, x_n$ образует подполе поля Ω. Это подполе мы обозначим через Δ и будем называть *полем алгебраических дробей* от $x_1, \ldots, x_n$ над P.

Мы рассматривали $P[x_1, \ldots, x_n]$ как подкольцо некоторого поля Ω. Не исключено, однако, что $P[x_1, \ldots, x_n]$ может оказаться подкольцом и другого поля Ω'. Возникает вопрос, каково будет поле алгебраических дробей в Ω'. Оказывается, что поле алгебраических дробей будет (с точностью до изоморфизма) тем же, что и в Ω. Иными словами, можно высказать следующую теорему:

Теорема 23. *Если для кольца многочленов $P[x_1, \ldots, x_n]$ существует поле алгебраических дробей, то это поле является единственным* (с точностью до изоморфизма).

Мы говорим «если существует», так как мы пока ещё не знаем, существует ли вообще поле Ω, имеющее $P[x_1, \ldots, x_n]$ своим подкольцом.

Доказательство. Пусть $P[x_1, \ldots, x_n]$ является подкольцом как поля Ω, так и поля Ω'. Обозначим через Δ поле алгебраических дробей в Ω и через Δ' поле алгебраических дробей в Ω'. Далее, возьмём из $P[x_1, \ldots, x_n]$ два произвольных многочлена $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n) \neq 0$ и обозначим через α решение уравнения

$$gz = f \tag{1}$$

в поле Δ и через α' решение того же уравнения в поле Δ'.

Поставим теперь в соответствие элементу α элемент α':

$$\alpha \to \alpha'. \tag{2}$$

Это соответствие является не только однозначным, но и взаимно однозначным. В самом деле, пусть решению β уравнения

$$g_1 z = f_1 \tag{3}$$

в поле Δ соответствует то же самое α'. Это значит, что уравнения (1) и (3) имеют в поле Δ' одно и то же решение α'. Отсюда в силу свойства 1° алгебраических дробей должно выполняться равенство

$$fg_1 = f_1 g. \tag{4}$$

Но если имеет место равенство (4), то в силу того же свойства 1° $\beta = \alpha$. Очевидно, что для каждого α' из Δ' найдётся в Δ элемент α, которому α' и ставится в соответствие. Всё это вместе взятое означает, что (2) есть взаимно однозначное соответствие между полями Δ и Δ'.

Покажем, что соответствие (2) является изоморфизмом между Δ и Δ'.

Пусть α и β — два каких-нибудь элемента из Δ, причём пусть α и β являются соответственно корнями уравнений

$$g_1 z = f_1 \qquad (g_1 \neq 0) \tag{5}$$

и

$$g_2 z = f_2 \qquad (g_2 \neq 0) \tag{6}$$

в поле Δ. Пусть в поле Δ' уравнение (5) имеет корень α', а уравнение (6) корень β'. Тогда

$$\alpha \to \alpha', \quad \beta \to \beta'.$$

Согласно свойству 2° алгебраических дробей сумма $\alpha + \beta$ есть корень уравнения

$$g_1 g_2 z = f_1 g_2 + f_2 g_1 \tag{7}$$

в поле Δ. Обозначим корень уравнения (7) в поле Δ' через γ'. Тогда

$$\alpha + \beta \to \gamma'.$$

Обращаясь снова к свойству 2°, имеем, что сумма $\alpha' + \beta'$ должна быть в поле Δ' корнем того же уравнения (7), откуда в силу единственности решения уравнения (7) $\gamma' = \alpha' + \beta'$. Мы видим, что

$$\alpha + \beta \to \alpha' + \beta'.$$

Подобным же образом, пользуясь свойством 3° алгебраических дробей, можно показать, что $\alpha\beta \to \alpha'\beta'$.

Итак, поля Δ и Δ' оказались изоморфными.

Отметим, что при изоморфизме (2) многочлены из $P[x_1, \ldots, x_n]$ будут соответствовать самим себе: $f \to f$. Действительно, многочлен $f(x_1, \ldots, x_n)$ можно рассматривать как корень уравнения

$$1 \cdot z = f,$$

где 1 — единица поля P. Но это уравнение как в поле Δ, так и в поле Δ' имеет одно и то же решение, а именно f.

Мы обнаружили, что если поле алгебраических дробей для кольца $P[x_1, \ldots, x_n]$ существует, то оно является единственным. Покажем теперь, что это поле в самом деле существует.

Теорема 24 (о существовании поля алгебраических дробей). *Поле алгебраических дробей существует для всякого кольца многочленов* $P[x_1, \ldots, x_n]$.

Доказательство. Рассмотрим множество M пар (f, g) многочленов $f = f(x_1, \ldots, x_n)$ и $g = g(x_1, \ldots, x_n) \neq 0$ из $P[x_1, \ldots, x_n]$. Руководствуясь свойством 1° алгебраических дробей, мы для пар (f, g) введём следующее отношение эквивалентности: положим $(f_1, g_1) \sim (f_2, g_2)$, если $f_1 g_2 = f_2 g_1$.

Легко убедиться, что определённое таким образом соотношение действительно является отношением эквивалентности между парами множества M.

В самом деле, так как $f_1g_1=f_1g_1$, то $(f_1, g_1)\sim(f_1, g_1)$, т. е. наше соотношение обладает свойством рефлексивности.

Далее, так как равенство $f_1g_2=f_2g_1$ можно переписать в виде $f_2g_1=f_1g_2$, то из $(f_1, g_1)\sim(f_2, g_2)$ следует $(f_2, g_2)\sim(f_1, g_1)$, т. е. соотношение обладает свойством симметрии.

Наконец, пусть $(f_1, g_1)\sim(f_2, g_2)$ и $(f_2, g_2)\sim(f_3, g_3)$. Тогда будут иметь место равенства

$$f_1g_2=f_2g_1 \tag{8}$$

и

$$f_2g_3=f_3g_2. \tag{9}$$

Умножим обе части равенства (8) на g_3 и равенства (9) на g_1. Получим:

$$f_1g_2g_3=f_2g_1g_3, \quad f_2g_1g_3=f_3g_1g_2,$$

откуда

$$f_1g_2g_3=f_3g_1g_2.$$

Сокращая последнее равенство на g_2, имеем: $f_1g_3=f_3g_1$, т. е. из $(f_1, g_1)\sim(f_2, g_2)$ и $(f_2, g_2)\sim(f_3, g_3)$ следует, что $(f_1, g_1)\sim(f_3, g_3)$. Соотношение, таким образом, обладает свойством транзитивности.

Итак, соотношение $\sim$ обладает всеми свойствами эквивалентности. Тем самым это соотношение определяет разбиение множества M на классы эквивалентных пар. Условимся класс, в котором лежит пара (f, g), обозначать символом $\frac{f}{g}$. Очевидно, что $\frac{f_1}{g_1}=\frac{f_2}{g_2}$ тогда и только тогда, когда $f_1g_2=f_2g_1$.

Введём для множества Δ всех этих классов $\frac{f}{g}$ операции сложения и умножения так, чтобы Δ стало полем относительно введённых операций.

Руководствуясь свойствами 2° и 3° алгебраических дробей, мы сумму и произведение классов определим следующим образом:

$$\frac{f_1}{g_1}+\frac{f_2}{g_2}=\frac{f_1g_2+f_2g_1}{g_1g_2}, \tag{10}$$

$$\frac{f_1}{g_1}\cdot\frac{f_2}{g_2}=\frac{f_1f_2}{g_1g_2}. \tag{11}$$

Покажем, что эти определения законны. Они законны, во-первых, потому, что $g_1g_2\neq 0$, так как $g_1\neq 0$ и $g_2\neq 0$ (напоминаем — для всех пар (f, g) множества M многочлен $g\neq 0$). Таким образом, символы в правых частях равенств (10) и (11) имеют смысл.

Во-вторых, правые части равенств (10) и (11) не зависят от выбора представителей классов. В самом деле, пусть

$$\frac{f_1}{g_1}=\frac{\psi}{\varphi}. \tag{12}$$

Посмотрим, что произойдёт, если в сумме $\frac{f_1}{g_1}+\frac{f_2}{g_2}$ и в произведении $\frac{f_1}{g_1}\cdot\frac{f_2}{g_2}$ заменить $\frac{f_1}{g_1}$ через $\frac{\psi}{\varphi}$.

Из равенства (12) следует, что

$$f_1\varphi=\psi g_1. \tag{13}$$

Умножая обе части равенства (13) на g_2, получаем:

$$f_1g_2\varphi=\psi g_1g_2.$$

Прибавляя затем к обеим частям последнего равенства $f_2g_1\varphi$, имеем:

$$f_1g_2\varphi+f_2g_1\varphi=\psi g_1g_2+f_2g_1\varphi,$$
$$(f_1g_2+f_2g_1)\varphi=(\psi g_2+f_2\varphi)g_1.$$

Наконец, умножаем обе части получившегося равенства на g_2:

$$(f_1g_2+f_2g_1)\varphi g_2=(\psi g_2+f_2\varphi)g_1g_2. \tag{14}$$

Равенство (14) как раз свидетельствует о том, что

$$\frac{f_1g_2+f_2g_1}{g_1g_2}=\frac{\psi g_2+f_2\varphi}{\varphi g_2}.$$

Аналогично, умножая обе части равенства (13) на f_2g_2, получаем:

$$f_1f_2\varphi g_2=\psi f_2g_1g_2,$$

откуда

$$\frac{f_1f_2}{g_1g_2}=\frac{\psi f_2}{\varphi g_2}.$$

Совершенно тем же способом можно показать, что, выбирая вместо (f_2, g_2) другой представитель класса $\frac{f_2}{g_2}$, мы этим также не изменим правые части равенств (10) и (11).

Теперь покажем, что относительно введённых операций множество классов Δ образует поле. Очевидно, придётся проверить все условия, характеризующие поле. Мы ограничимся проверкой сочетательности сложения. Имеем:

$$\frac{f_1}{g_1}+\left(\frac{f_2}{g_2}+\frac{f_3}{g_3}\right)=\frac{f_1}{g_1}+\frac{f_2g_3+f_3g_2}{g_2g_3}=\frac{f_1g_2g_3+f_2g_1g_3+f_3g_1g_2}{g_1g_2g_3};$$

$$\left(\frac{f_1}{g_1}+\frac{f_2}{g_2}\right)+\frac{f_3}{g_3}=\frac{f_1g_2+f_2g_1}{g_1g_2}+\frac{f_3}{g_3}=\frac{f_1g_2g_3+f_2g_1g_3+f_3g_1g_2}{g_1g_2g_3}.$$

Получился в обоих случаях один и тот же результат. Следовательно,

$$\frac{f_1}{g_1}+\left(\frac{f_2}{g_2}+\frac{f_3}{g_3}\right)=\left(\frac{f_1}{g_1}+\frac{f_2}{g_2}\right)+\frac{f_3}{g_3}.$$

Так же проверяются и остальные условия.

Мы утверждаем, что построенное поле Δ и будет искомым полем алгебраических дробей. Это утверждение станет очевидным, если мы покажем, что Δ содержит подкольцо, изоморфное кольцу $P[x_1, \ldots, x_n]$.

Рассмотрим в поле Δ подмножество классов вида $\frac{f}{1}$, где 1 — единица поля P. Поставим в соответствие каждому многочлену f из $P[x_1, \ldots, x_n]$ класс $\frac{f}{1}$ с тем же самым f. Мы получим тогда взаимно однозначное соответствие

$$f \to \frac{f}{1} \tag{15}$$

между кольцом $P[x_1, \ldots, x_n]$ и множеством, состоящим из классов вида $\frac{f}{1}$.

В самом деле, разным многочленам f_1 и f_2 из $P[x_1, \ldots, x_n]$ должны соответствовать разные классы $\frac{f_1}{1} \neq \frac{f_2}{1}$. Если бы классы $\frac{f_1}{1}$ и $\frac{f_2}{1}$ совпадали, то на основании условия равенства двух классов мы имели бы, что $f_1 \cdot 1 = f_2 \cdot 1$ или $f_1 = f_2$, что невозможно.

Взаимно однозначное соответствие (15) будет к тому же изоморфизмом. Действительно, если

$$f_1 \to \frac{f_1}{1}, \quad f_2 \to \frac{f_2}{1},$$

то

$$f_1+f_2 \to \frac{f_1+f_2}{1}=\frac{f_1\cdot 1+f_2\cdot 1}{1\cdot 1}=\frac{f_1}{1}+\frac{f_2}{1},$$

$$f_1 f_2 \to \frac{f_1 f_2}{1}=\frac{f_1 f_2}{1\cdot 1}=\frac{f_1}{1}\cdot\frac{f_2}{1}.$$

Итак, множество классов вида $\frac{f}{1}$ образует подкольцо поля Δ, изоморфное кольцу $P[x_1, \ldots, x_n]$. В силу этого мы можем класс $\frac{f}{1}$ не отличать от соответствующего многочлена f.

Далее, так как класс $\frac{1}{g}$ является элементом, обратным относительно класса $\frac{g}{1}$, то в силу равенства

$$\frac{f}{1}\cdot\frac{1}{g}=\frac{f}{g}$$

мы можем любой классс $\frac{f}{g}$, т. е. любой элемент поля Δ, рассматривать как отношение многочленов f и $g \neq 0$ из кольца $P[x_1, \ldots, x_n]$, т. е. Δ есть поле алгебраических дробей для кольца $P[x_1, \ldots, x_n]$. Теорема доказана.

Поскольку мы убедились в существовании поля алгебраических дробей для кольца $P[x_1, \ldots, x_n]$, мы будем в дальнейшем обозначать это поле через $P(x_1, \ldots, x_n)$ (неизвестные в круглых, а не в квадратных скобках).

Перейдём теперь к вопросу о функциональной точке зрения на алгебраическую дробь.

Обратимся к случаю алгебраической дроби от одного неизвестного.

Пусть

$$r(x) = \frac{f(x)}{g(x)} \tag{16}$$

— некоторая алгебраическая дробь из $P(x)$. Введём понятие значения дроби $r(x)$.

Предварительно отметим, что числитель и знаменатель дроби (16) можно всегда предполагать взаимно простыми. Действительно, если бы многочлены $f(x)$ и $g(x)$ имели наибольший общий делитель $D(x)$ со степенью выше нулевой, то

$$f(x) = f_1(x) D(x), \quad g(x) = g_1(x) D(x),$$

$f_1(x)$ и $g_1(x)$ были бы взаимно простыми, и на основании условия равенства алгебраических дробей мы могли бы написать:

$$r(x) = \frac{f(x)}{g(x)} = \frac{f_1(x)}{g_1(x)}.$$

Итак, мы будем многочлены $f(x)$ и $g(x)$ предполагать взаимно простыми. Возьмём теперь из поля P некоторый элемент c. Если $g(c) \neq 0$, то под значением $r(x)$ при $x = c$ мы будем подразумевать отношение $\frac{f(c)}{g(c)}$ значений многочленов $f(x)$ и $g(x)$ при $x = c$ и будем это отношение обозначать через $r(c)$. Очевидно, что значение $r(c)$ дроби $r(x)$ есть некоторый элемент поля P.

Из этого определения следует, что *если две алгебраические дроби равны: $r_1(x) = r_2(x)$, то их значения совпадают при любом значении неизвестного x, не обращающего в нуль знаменатели дробей $r_1(x)$ и $r_2(x)$.*

В самом деле, если

$$r_1(x) = \frac{f_1(x)}{g_1(x)}, \quad r_2(x) = \frac{f_2(x)}{g_2(x)}$$

и $r_1(x) = r_2(x)$, то в силу условия равенства алгебраических дробей

$$f_1(x) g_2(x) = f_2(x) g_1(x).$$

Пусть c — элемент из P, для которого $g_1(c) \neq 0$ и $g_2(c) \neq 0$. Полагая $x = c$, имеем:

$$f_1(c)\, g_2(c) = f_2(c)\, g_1(c),$$

откуда, пользуясь условием равенства дробей в поле P, получаем:

$$\frac{f_1(c)}{g_1(c)} = \frac{f_2(c)}{g_2(c)}.$$

Покажем, что в случае бесконечного поля P будет верно и обратное: *если значения алгебраических дробей $r_1(x)$ и $r_2(x)$ совпадают при любом значении неизвестного x, не обращающем в нуль знаменатели дробей, то дроби $r_1(x)$ и $r_2(x)$ равны.*

Доказательство. Пусть

$$\frac{f_1(c)}{g_1(c)} = \frac{f_2(c)}{g_2(c)},$$

где c — произвольный элемент из P, не обращающий в нуль $g_1(x)$ и $g_2(x)$. Тогда

$$f_1(c)\, g_2(c) = f_2(c)\, g_1(c). \tag{17}$$

Из равенства (17) видно, что значения произведений $f_1(x)\, g_2(x)$ и $f_2(x)\, g_1(x)$ совпадают при бесконечном множестве значений неизвестного x, так как поле P бесконечно, а многочлены $g_1(x) \neq 0$ и $g_2(x) \neq 0$ над полем P имеют ограниченное число корней. Следовательно, многочлены $f_1(x)\, g_2(x)$ и $f_2(x)\, g_1(x)$ должны быть равны

$$f_1(x)\, g_2(x) = f_2(x)\, g_1(x).$$

Отсюда на основании равенства алгебраических дробей получается, что

$$\frac{f_1(x)}{g_1(x)} = \frac{f_2(x)}{g_2(x)},$$

и наше предложение доказано.

Будем в произвольной алгебраической дроби $r(x) = \frac{f(x)}{g(x)}$ из $P(x)$ неизвестное x заменять тем или иным элементом c из поля P, не обращающим в нуль знаменатель $g(x)$. Тогда мы будем получать вполне определённые элементы $r(c)$ из P. Таким образом, каждой алгебраической дроби $r(x)$ из $P(x)$ будет ставиться в соответствие функция от одного аргумента ξ, определённая для всех значений x, кроме значений, обращающих в нуль знаменатель $g(x)$:

$$r(x) \to r(\xi). \tag{18}$$

Через $r(\xi)$ мы здесь обозначили функцию, соответствующую алгебраической дроби $r(x)$. Мы будем $r(\xi)$ называть *рациональной функцией* над полем P.

Руководствуясь соответствием (18), введём теперь операции сложения и умножения рациональных функций. А именно, если

$$r_1(x)=\frac{f_1(x)}{g_1(x)}, \quad r_2(x)=\frac{f_2(x)}{g_2(x)}$$

— две какие-нибудь алгебраические дроби из $P(x)$, то под $r_1(\xi)+r_2(\xi)$ мы будем подразумевать рациональную функцию, соответствующую сумме $r_1(x)+r_2(x)$ алгебраических дробей, а под $r_1(\xi)\,r_2(\xi)$ будем подразумевать рациональную функцию, соответствующую произведению $r_1(x)\,r_2(x)$ алгебраических дробей.

Наше определение операций сложения и умножения несколько расходится с обычным определением операций над функциями. Так, например, в нашем смысле

$$(\xi-1)^2\cdot\frac{1}{\xi-1}=\xi-1 \qquad (1 \text{ — единица поля } P), \tag{19}$$

а в обычном смысле $(\xi-1)^2\cdot\dfrac{1}{\xi-1}$ нельзя считать равным $\xi-1$, так как правая часть равенства (19) определена для всех значений ξ, а в левой части $\dfrac{1}{\xi-1}$ при $\xi=1$ не определена.

Итак, мы определили надлежащим образом операции сложения и умножения рациональных функций над полем P. Покажем теперь, что в случае бесконечного поля P алгебраическая и функциональная точки зрения на дробь в известном смысле слова совпадают. Точнее, покажем, что имеет место следующая

Теорема 25. *Если поле P бесконечно, то множество рациональных функций $r(\xi)$ над P образует поле, изоморфное полю алгебраических дробей $P(x)$.*

Доказательство. Пусть двум алгебраическим дробям

$$r_1(x)=\frac{f_1(x)}{g_1(x)} \quad \text{и} \quad r_2(x)=\frac{f_2(x)}{g_2(x)}$$

соответствует одна и та же рациональная функция $r(\xi)$ над P:

$$r_1(x)\to r(\xi), \quad r_2(x)\to r(\xi).$$

Тогда $r_1(c)=r_2(c)$ для любого c из P, не обращающего в нуль знаменатели $g_1(x)$ и $g_2(x)$ дробей $r_1(x)$ и $r_2(x)$. Но выше мы уже убедились, что в случае бесконечного поля P такие алгебраические дроби должны быть равны. Следовательно, $r_1(x)=r_2(x)$, и мы видим, что соответствие (18) является не только однозначным, но и взаимно однозначным.

Далее, согласно самому определению суммы и произведения рациональных функций

$$r_1(x)+r_2(x)\to r_1(\xi)+r_2(\xi),$$
$$r_1(x)\,r_2(x)\to r_1(\xi)\,r_2(\xi).$$

Таким образом, соответствие (18) есть изоморфизм между полем $P(x)$ и множеством рациональных функций $r(\xi)$ над P. Тем самым это множество функций образует поле, изоморфное $P(x)$, и теорема доказана.

Только что доказанная теорема даёт нам основание не различать в случае бесконечного поля P алгебраическую дробь от рациональной функции [1]), и мы можем в этом случае аргумент ξ обозначать той же буквой, что и неизвестное x.

Примерно такое же обоснование функциональной точки зрения на алгебраическую дробь можно дать и для случая нескольких неизвестных, но мы не будем на этом задерживаться.

§ 13. Симметрические многочлены

В настоящем параграфе будет рассмотрен один довольно важный класс алгебраических дробей, так называемые симметрические алгебраические дроби или симметрические функции.

Пусть R — коммутативное кольцо с единицей $e \neq 0$ и без делителей нуля. *Симметрическим многочленом* над кольцом R принято называть такой многочлен $f(x_1, \ldots, x_n)$ от n неизвестных над R, который не меняется при любой перестановке неизвестных $x_1, \ldots, x_n$.

Например, многочлен

$$f(x_1, x_2, x_3) = x_1^2x_2 + x_1x_2^2 + x_1^2x_3 + x_1x_3^2 + x_2^2x_3 + x_2x_3^2 \qquad (1)$$

является симметрическим; легко убедиться, что он не меняется при любой перестановке неизвестных:

$$f(x_1, x_2, x_3) = f(x_3, x_2, x_1) = f(x_2, x_1, x_3) = f(x_2, x_3, x_1) = \\ = f(x_1, x_3, x_2) = f(x_3, x_1, x_2).$$

Возьмём хотя бы

$$f(x_3, x_2, x_1) = x_3^2x_2 + x_3x_2^2 + x_3^2x_1 + x_3x_1^2 + x_2^2x_1 + x_2x_1^2. \qquad (2)$$

Для получения выражения $f(x_3, x_2, x_1)$ мы в выражении (1) многочлена $f(x_1, x_2, x_3)$ неизвестное x_1 заменили через x_3, x_2 оставили без изменения, а x_3 заменили через x_1. Сравнивая выражения (1) и (2), видим, что они отличаются друг от друга лишь порядком следования членов и порядком следования неизвестных в каждом члене. Следовательно, $f(x_3, x_2, x_1) = f(x_1, x_2, x_3)$.

Алгебраическая дробь

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)},$$

у которой числитель $f(x_1, \ldots, x_n)$ и знаменатель $g(x_1, \ldots, x_n) \neq 0$ являются симметрическими многочленами над полем P, называется *симметрической алгебраической дробью*, или *симметрической функцией* над P.

[1]) Не различать по отношению к алгебраическим операциям.

Поскольку симметрическая функция выражается через симметрические многочлены, мы, естественно, начнём наше изложение с изучения основных свойств симметрических многочленов.

Впервые приходится сталкиваться с симметрическими многочленами при решении следующей задачи: пусть уравнение n-й степени над полем P

$$f(x) = x^n + a_1 x^{n-1} + \ldots + a_n = 0 \tag{3}$$

со старшим коэффициентом, равным единице, имеет n корней. Эти корни $\alpha_1, \alpha_2, \ldots, \alpha_n$ даны. Найти коэффициенты a_i уравнения.

Мы знаем (см. § 8, стр. 208), что коэффициенты уравнения (3) выражаются через корни по формулам Вьета:

$$\left.\begin{aligned} a_1 &= -(\alpha_1 + \alpha_2 + \ldots + \alpha_n), \\ a_2 &= (\alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_{n-1}\alpha_n), \\ a_3 &= -(\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_3\alpha_4 + \ldots + \alpha_{n-2}\alpha_{n-1}\alpha_n), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_n &= (-1)^n \alpha_1\alpha_2 \ldots \alpha_n. \end{aligned}\right\} \tag{4}$$

Руководствуясь этими формулами, составим теперь многочлены от n неизвестных

$$\left.\begin{aligned} \sigma_1 &= x_1 + x_2 + \ldots + x_n, \\ \sigma_2 &= x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \sigma_n &= x_1x_2 \ldots x_n. \end{aligned}\right\} \tag{5}$$

Легко видеть, что многочлены (5) являются симметрическими. В самом деле, равенства (4), очевидно, не зависят от нумерации корней: $\alpha_1, \alpha_2, \ldots, \alpha_n$. Мы могли бы, например, корню α_1 приписать другой номер, хотя бы 2, а корню α_3 — номер, равный единице; это изменение нумерации ни в какой мере не нарушило бы равенств (4), так как при их выводе совершенно безразлично, какой корень следует обозначать через α_1, какой через α_2 и т. д.

Многочлены (5) называются *основными* или *элементарными симметрическими многочленами* от неизвестных $x_1, \ldots, x_n$.

Эти многочлены играют в теории симметрических функций исключительную роль благодаря следующей теореме:

Основная теорема о симметрических многочленах. *Всякий симметрический многочлен $f(x_1, \ldots, x_n)$ от n неизвестных над кольцом R может быть выражен в виде многочлена от основных симметрических многочленов $\sigma_1, \sigma_2, \ldots, \sigma_n$ над тем же кольцом R:*

$$f(x_1, \ldots, x_n) = g(\sigma_1, \ldots, \sigma_n),$$

где $g(\sigma_1, \ldots, \sigma_n)$ — многочлен от $\sigma_1, \sigma_2, \ldots, \sigma_n$ над R.

Существует весьма большое количество доказательств этой теоремы. Мы предлагаем вниманию читателя доказательство, достаточно простое как в теоретическом, так и в практическом отношении. Однако предварительно придётся ввести понятие высоты члена многочлена и доказать одну лемму, связанную с этим понятием.

Пусть $f(x_1, \ldots, x_n)$ — некоторый многочлен (он может быть и несимметрическим). Возьмём два каких-нибудь члена этого многочлена

$$Ax_1^{\alpha_1}x_2^{\alpha_2} \ldots x_n^{\alpha_n} \tag{6}$$

и

$$Bx_1^{\beta_1}x_2^{\beta_2} \ldots x_n^{\beta_n}. \tag{7}$$

Условимся считать из членов (6) и (7) тот выше, у которого больше показатель при x_1; если же показатели при x_1 равны, то условимся считать выше тот член, у которого больше показатель при x_2, и т. д. Вообще, если $\alpha_i - \beta_i$ есть первая отличная от нуля разность, то при $\alpha_i - \beta_i > 0$ член (6) считается выше члена (7), а при $\alpha_i - \beta_i < 0$ — ниже члена (7).

Пример 1. Какой из членов многочлена

$$f(x_1, x_2, x_3) = 8x_1x_2x_3^3 + x_1^4x_2 + x_1^4x_2^2x_3 - x_1^4x_2^2x_3^2$$

является наивысшим?

Член $8x_1x_2x_3^3$ ниже члена $x_1^4x_2$; в свою очередь член $x_1^4x_2$ ниже члена $x_1^4x_2^2x_3$, а $x_1^4x_2^2x_3$ ниже члена $-x_1^4x_2^2x_3^2$. Таким образом, данный многочлен имеет наивысшим членом $-x_1^4x_2^2x_3^2$.

Мы предостерегаем читателя от смешения понятий высшего члена и члена с наибольшей степенью. Так, в только что разобранном примере член $8x_1x_2x_3^3$ имеет наибольшую степень, но он является даже наинизшим членом многочлена.

Лемма. *Наивысший член произведения двух многочленов* $f(x_1, \ldots, x_n)$ *и* $g(x_1, \ldots, x_n)$ *равен произведению наивысших членов этих многочленов.*

Доказательство. Пусть

$$Ax_1^{\alpha_1}x_2^{\alpha_2} \ldots x_n^{\alpha_n} \tag{8}$$

и

$$Bx_1^{\beta_1}x_2^{\beta_2} \ldots x_n^{\beta_n} \tag{9}$$

— соответственно наивысшие члены многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$. Возьмём какой-нибудь член

$$Mx_1^{\mu_1}x_2^{\mu_2} \ldots x_n^{\mu_n} \tag{10}$$

многочлена $f(x_1, \ldots, x_n)$ и какой-нибудь член

$$Nx_1^{\lambda_1}x_2^{\lambda_2} \ldots x_n^{\lambda_n} \tag{11}$$

многочлена $g(x_1, \ldots, x_n)$ и покажем, что произведение

$$ABx_1^{\alpha_1+\beta_1}x_2^{\alpha_2+\beta_2} \ldots x_n^{\alpha_n+\beta_n} \tag{12}$$

членов (8) и (9) выше произведения

$$MNx_1^{\mu_1+\lambda_1}x_2^{\mu_2+\lambda_2} \ldots x_n^{\mu_n+\lambda_n} \tag{13}$$

членов (10) и (11), если член (10) ниже члена (8) или член (11) ниже члена (9).

Так как (8) и (9) — наивысшие члены многочленов $f(x_1, \ldots, x_n)$ и $g(x_1, \ldots, x_n)$, то $\alpha_1 \geqslant \mu_1$, $\beta_1 \geqslant \lambda_1$, откуда $\alpha_1 + \beta_1 \geqslant \mu_1 + \lambda_1$. Таким образом, если $\alpha_1 + \beta_1 > \mu_1 + \lambda_1$, то член (12) выше члена (13). Если же $\alpha_1 + \beta_1 = \mu_1 + \lambda_1$, то $\alpha_1 = \mu_1$, $\beta_1 = \lambda_1$, и мы обращаемся к следующим показателям α_2 и β_2. Так как $\alpha_1 = \mu_1$, $\beta_1 = \lambda_1$ и член (8) не ниже члена (10), а член (9) не ниже члена (11), то $\alpha_2 \geqslant \mu_2$, $\beta_2 \geqslant \lambda_2$, откуда $\alpha_2 + \beta_2 \geqslant \mu_2 + \lambda_2$. Член (12) будет выше члена (13), если $\alpha_2 + \beta_2 > \mu_2 + \lambda_2$; если же $\alpha_2 + \beta_2 = \mu_2 + \lambda_2$, то $\alpha_2 = \mu_2$, $\beta_2 = \lambda_2$, и мы обращаемся к показателям α_3 и β_3 и т. д. В конечном счёте мы доберёмся до показателей α_k и β_k таких, что будет иметь место по меньшей мере одно из неравенств $\alpha_k > \mu_k$ и $\beta_k > \lambda_k$, и тогда мы получим, что $\alpha_k + \beta_k > \mu_k + \lambda_k$, т. е. член (12) выше члена (13).

Теперь приступаем к доказательству основной теоремы теории симметрических функций.

Доказательство. Возьмём наивысший член симметрического многочлена $f(x_1, \ldots, x_n)$. Пусть это будет член

$$Ax_1^{\alpha_1}x_2^{\alpha_2} \ldots x_n^{\alpha_n}. \tag{14}$$

Тогда многочлен $f(x_1, \ldots, x_n)$, являясь симметрическим, должен вместе с членом (14) содержать и все члены, получающиеся из члена (14) при всевозможных перестановках неизвестных. Покажем, что

$$\alpha_1 \geqslant \alpha_2 \geqslant \ldots \geqslant \alpha_n. \tag{15}$$

Действительно, переставляя x_1 и x_2, мы получаем из члена (14) член

$$Ax_1^{\alpha_2}x_2^{\alpha_1} \ldots x_n^{\alpha_n}.$$

Так как этот член не может быть выше члена (14), то показатель при x_1 в этом члене не может превосходить показателя при x_1 в члене (14): $\alpha_1 \geqslant \alpha_2$.

Точно так же, сравнивая член (14) с членом

$$Ax_1^{\alpha_1}x_2^{\alpha_3} \ldots x_n^{\alpha_n},$$

получающимся из (14) путём перестановки неизвестных x_2 и x_3, приходим к заключению, что $\alpha_2 \geqslant \alpha_3$ и т. д.

Очевидно, что

$$A\sigma_1^{k_1}\sigma_2^{k_2} \ldots \sigma_n^{k_n}, \tag{16}$$

где k_i — целые неотрицательные числа, есть также симметрический многочлен от $x_1, \ldots, x_n$. Попытаемся подобрать числа k_i так, чтобы наивысший член симметрического многочлена (16) совпадал с членом (14).

Основные симметрические многочлены $\sigma_1, \sigma_2, \ldots, \sigma_n$ имеют наивысшими членами соответственно $x_1, x_1x_2, \ldots, x_1x_2 \ldots x_n$. Следовательно, согласно недавно доказанной лемме, симметрический многочлен (16) имеет наивысшим членом

$$Ax_1^{k_1}(x_1x_2)^{k_2} \ldots (x_1x_2 \ldots x_n)^{k_n} = \\ = Ax_1^{k_1+k_2+\ldots+k_n}x_2^{k_2+\ldots+k_n} \ldots x_n^{k_n}. \quad (17)$$

Таким образом, член (17) будет совпадать с членом (14) в том случае, если

$$\begin{aligned} k_1 + k_2 + \ldots + k_n &= \alpha_1, \\ k_2 + \ldots + k_n &= \alpha_2, \\ \ldots\ldots\ldots\ldots & \\ k_n &= \alpha_n. \end{aligned}$$

Решая эту систему уравнений относительно k_i, получаем:

$$k_1 = \alpha_1 - \alpha_2, \quad k_2 = \alpha_2 - \alpha_3, \quad \ldots, \quad k_{n-1} = \alpha_{n-1} - \alpha_n, \quad k_n = \alpha_n.$$

В силу неравенств (15) эти значения k_i будут целыми неотрицательными числами.

Итак, вычитая из многочлена $f(x_1, \ldots, x_n)$ выражение

$$A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3} \ldots \sigma_{n-1}^{\alpha_n-1-\alpha_n}\sigma_n^{\alpha_n},$$

мы уничтожим член (14) и получим симметрический многочлен

$$f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3} \ldots \sigma_n^{\alpha_n},$$

состоящий из более низких членов. Пусть

$$Bx_1^{\beta_1}x_2^{\beta_2} \ldots x_n^{\beta_n}$$

— наивысший член многочлена $f_1(x_1, \ldots, x_n)$. Тогда мы снова повторим процесс понижения высоты членов — вычтем из многочлена $f_1(x_1, \ldots, x_n)$ выражение $B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3} \ldots \sigma_n^{\beta_n}$, в результате чего получим симметрический многочлен

$$f_2(x_1, \ldots, x_n) = f_1(x_1, \ldots, x_n) - B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3} \ldots \sigma_n^{\beta_n}$$

и т. д. Этот процесс, однако, не может быть бесконечным: если на k-м шагу мы придём к симметрическому многочлену $f_k(x_1, \ldots, x_n)$ с наивысшим членом

$$Lx_1^{\lambda_1}x_2^{\lambda_2} \ldots x_n^{\lambda_n}, \quad (18)$$

то, с одной стороны, его показатели λ_i будут удовлетворять условию $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$, а с другой стороны, $\alpha_1 \geqslant \lambda_1$, так как член

(14) выше члена (18). Но, очевидно, существует лишь конечное множество систем целых неотрицательных чисел $\lambda_1, \lambda_2, \ldots, \lambda_n$, удовлетворяющих условиям $\alpha_1 \geqslant \lambda_1$ и $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$. Тем самым наш процесс должен закончиться, т. е. через конечное число шагов неизбежно получится, что

$$f_{s-1}(x_1, \ldots, x_n) - H\sigma_1^{\omega_1-\omega_2}\sigma_2^{\omega_2-\omega_3} \ldots \sigma_n^{\omega_n} = 0.$$

Отсюда вытекает, что

$$f(x_1, \ldots, x_n) = A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3} \ldots \sigma_n^{\alpha_n} + \\ + B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3} \ldots \sigma_n^{\beta_n} + \ldots + H\sigma_1^{\omega_1-\omega_2}\sigma_2^{\omega_2-\omega_3} \ldots \sigma_n^{\omega_n},$$

т. е. многочлен $f(x_1, \ldots, x_n)$ выразился в виде многочлена от σ_1, $\sigma_2, \ldots, \sigma_n$ над тем же самым кольцом R. Этим теорема и доказана полностью.

Способ выражения симметрического многочлена через основные симметрические многочлены, положенный в основу только что проведённого доказательства, является и в практическом отношении довольно удобным способом. Обратимся к такому примеру.

Пример 2. Выразить симметрический многочлен

$$f(x_1, x_2, x_3) = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2 \qquad (19)$$

через основные симметрические многочлены σ_1, σ_2, σ_3.

Очевидно, что $x_1 - x_2$, $x_1 - x_3$, $x_2 - x_3$ имеют наивысшими членами соответственно x_1, x_1 и x_2. Поэтому на основании известной леммы наивысшим членом многочлена (19) будет:

$$x_1^2 x_1^2 x_2^2 = x_1^4 x_2^2.$$

Существенно облегчает наши выкладки то обстоятельство, что многочлен (19) является формой шестой степени.

Составим теперь следующую табличку наивысших членов многочленов $f(x_1, \ldots, x_n)$, $f_1(x_1, \ldots, x_n)$, $f_2(x_1, \ldots, x_n)$ и т. д., получающихся при нашем способе понижения высоты членов симметрического многочлена. Ясно, что каждый из этих многочленов будет также формой шестой степени, и поэтому все члены будут шестой степени:

Система показателей наивысшего члена	Наивысший член	Соответствующая комбинация основных симметрических функций
4 2 0	$x_1^4 x_2^2$	$\sigma_1^{4-2}\sigma_2^{2-0}\sigma_3^0 = \sigma_1^2\sigma_2^2$
4 1 1	$Bx_1^4 x_2 x_3$	$B\sigma_1^{4-1}\sigma_2^{1-1}\sigma_3 = B\sigma_1^3\sigma_3$
3 3 0	$Cx_1^3 x_2^3$	$C\sigma_1^{3-3}\sigma_2^{3-0}\sigma_3^0 = C\sigma_2^3$
3 2 1	$Dx_1^3 x_2^2 x_3$	$D\sigma_1^{3-2}\sigma_2^{2-1}\sigma_3 = D\sigma_1\sigma_2\sigma_3$
2 2 2	$Ex_1^2 x_2^2 x_3^2$	$E\sigma_1^{2-2}\sigma_2^{2-2}\sigma_3^2 = E\sigma_3^2$

Обращаем внимание читателя на то, что при составлении первого столбца таблички следует руководствоваться условием $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$, имеющим место для наивысшего члена симметрического многочлена $f_k(x_1, \ldots, x_n)$, получающегося на k-м шагу нашего процесса.

Таким образом,

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 + B\sigma_1^3\sigma_3 + C\sigma_2^3 + D\sigma_1\sigma_2\sigma_3 + E\sigma_3^2. \qquad (20)$$

Остаётся определить, чему равны коэффициенты B, C, D, E. Для этой цели будем неизвестным x_1, x_2, x_3 давать те или иные значения.

Полагаем $x_1 = 1$, $x_2 = 1$, $x_3 = 0$. Тогда

$$f(1, 1, 0) = 0, \quad \sigma_1 = 2, \quad \sigma_2 = 1, \quad \sigma_3 = 0,$$

и равенство (20) превращается в

$$0 = 4 + C,$$

откуда $C = -4$.

Следовательно,

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 + B\sigma_1^3\sigma_3 - 4\sigma_2^3 + D\sigma_1\sigma_2\sigma_3 + E\sigma_3^2. \qquad (21)$$

Полагаем теперь $x_1 = 1$, $x_2 = 1$, $x_3 = -2$. Получаем:

$$f(1, 1, -2) = 0, \quad \sigma_1 = 0, \quad \sigma_2 = -3, \quad \sigma_3 = -2,$$

и равенство (21) превращается в

$$0 = 108 + 4E,$$

откуда $E = -27$. Следовательно,

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 + B\sigma_1^3\sigma_3 - 4\sigma_2^3 + D\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2.$$

Далее, полагаем $x_1 = 1$, $x_2 = 1$, $x_3 = 1$. Получаем:

$$f(1, 1, 1) = 0, \quad \sigma_1 = 3, \quad \sigma_2 = 3, \quad \sigma_3 = 1$$

и

$$0 = 81 + 27B - 108 + 9D - 27,$$

или

$$3B + D = 6. \qquad (22)$$

Затем, полагая $x_1 = 1$, $x_2 = 1$, $x_3 = -1$, получаем:

$$f(1, 1, -1) = 0, \quad \sigma_1 = 1, \quad \sigma_2 = -1, \quad \sigma_3 = -1$$

и

$$0 = 1 - B + 4 + D - 27,$$

или

$$B - D = -22. \qquad (23)$$

Решая систему уравнений (22) и (23), находим без труда, что

$$B = -4, \quad D = 18.$$

Итак, окончательно

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 - 4\sigma_1^3\sigma_3 - 4\sigma_2^3 + 18\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2.$$

В случае неоднородного симметрического многочлена мы рекомендуем разбить многочлен на сумму однородных симметрических многочленов и каждый из таких многочленов выражать через основные симметрические многочлены так, как это было сделано в только что рассмотренном примере.

Существует, впрочем, и много других способов выражения симметрических многочленов через основные симметрические многочлены; некоторые из этих способов читатель найдёт в книге А. К. Сушкевич, Основы высшей алгебры, изд. 4-е, Гостехиздат, 1941.

Несмотря на большое разнообразие способов выражения симметрических многочленов через основные симметрические многочлены, имеет место следующая

Теорема о единственности выражения симметрических многочленов через основные симметрические многочлены. *Всякий симметрический многочлен выражается единственным образом в виде многочлена от основных симметрических многочленов.*

Мы опускаем доказательство этой теоремы [1]).

Очевидно, что все основные результаты настоящего параграфа остаются в силе и для симметрической алгебраической дроби. А именно, всякая симметрическая алгебраическая дробь

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}$$

от n неизвестных над полем P может быть выражена в виде алгебраической дроби от основных симметрических многочленов σ_1, σ_2, σ_3, ..., σ_n над тем же полем P:

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} = \frac{\varphi(\sigma_1, \ldots, \sigma_n)}{\psi(\sigma_1, \ldots, \sigma_n)},$$

где $\varphi(\sigma_1, \ldots, \sigma_n)$ и $\psi(\sigma_1, \ldots, \sigma_n)$ — многочлены от $\sigma_1, \ldots, \sigma_n$ над P.

§ 14. Некоторые приложения теории симметрических многочленов

Ещё в элементарной алгебре приходится встречаться с задачей уничтожения иррациональности в знаменателе. Здесь мы рассмотрим эту задачу в полном объёме для случая числового поля.

---

[1]) Её доказательство можно найти в учебниках по высшей алгебре, например в учебнике Куроша [1] или в учебнике Окунева [3].

Пусть

$$\frac{f(x)}{g(x)} \qquad (g(x) \neq 0)$$

— некоторая алгебраическая дробь от x над числовым полем P, $\varphi(x)$ — некоторый многочлен n-й степени над полем P и $\theta_1, \ldots, \theta_n$ — комплексные корни многочлена $\varphi(x)$, причём пусть $\theta_1, \ldots, \theta_n$ не являются корнями $g(x)$. Задача уничтожения иррациональности в знаменателе заключается в следующем: требуется преобразовать дробно-рациональное выражение

$$\frac{f(\theta_1)}{g(\theta_1)} \tag{1}$$

так, чтобы оно оказалось равным целому рациональному выражению от θ_1 с коэффициентами из того же поля P:

$$\frac{f(\theta_1)}{g(\theta_1)} = h(\theta_1),$$

где $h(x)$ — некоторый многочлен над P.

Мы предлагаем вниманию читателя два решения этой задачи.

1) Умножим числитель и знаменатель дроби (1) на $g(\theta_2)\ldots g(\theta_n)$. Получим:

$$\frac{f(\theta_1)}{g(\theta_1)} = \frac{f(\theta_1)\,g(\theta_2)\ldots g(\theta_n)}{g(\theta_1)\,g(\theta_2)\ldots g(\theta_n)}.$$

Мы видим, что в качестве знаменателя получился симметрический многочлен $F(\theta_1, \ldots, \theta_n) = g(\theta_1)\,g(\theta_2)\ldots g(\theta_n)$ от $\theta_1, \ldots, \theta_n$. Следовательно, согласно основной теореме теории симметрических многочленов $F(\theta_1, \ldots, \theta_n)$ можно выразить в виде многочлена над P от основных симметрических многочленов $\sigma_1, \sigma_2, \ldots, \sigma_n$. Отсюда в силу формул Вьета $F(\theta_1, \ldots, \theta_n)$ будет выражаться и через коэффициенты многочлена $\varphi(x)$, т. е. если

$$\varphi(x) = x^n + a_1 x^{n-1} + \ldots + a_n,$$

то $F(\theta_1, \ldots, \theta_n) = H(a_1, \ldots, a_n)$, где $H(a_1, \ldots, a_n)$ — многочлен от $a_1, a_2, \ldots, a_n$ над полем P. Но $a_1, a_2, \ldots, a_n$ суть числа из поля P. Значит, и $F(\theta_1, \ldots, \theta_n) = H(a_1, \ldots, a_n)$ является некоторым числом b из P. Таким образом,

$$\frac{f(\theta_1)}{g(\theta_1)} = \frac{1}{b} f(\theta_1)\,g(\theta_2)\ldots g(\theta_n).$$

Остаётся $f(\theta_1)\,g(\theta_2)\ldots g(\theta_n)$ выразить через θ_1. Для этой цели рассмотрим: $g(\theta_2)\,g(\theta_3)\ldots g(\theta_n)$. Произведение

$$g(\theta_2)\,g(\theta_3)\ldots g(\theta_n) \tag{2}$$

является, очевидно, симметрическим многочленом от $\theta_2, \ldots, \theta_n$. Следовательно, произведение (2) можно выразить через основные симметрические многочлены

$$\begin{aligned}
\bar{\sigma}_1 &= \theta_2 + \theta_3 + \ldots + \theta_n,\\
\bar{\sigma}_2 &= \theta_2\theta_3 + \ldots + \theta_{n-1}\theta_n,\\
&\ldots\ldots\ldots\ldots\\
\bar{\sigma}_{n-1} &= \theta_2\theta_3 \ldots \theta_n
\end{aligned}$$

от $\theta_2, \ldots, \theta_n$. В свою очередь $\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_{n-1}$ можно следующим образом выразить через θ_1 и основные симметрические многочлены $\sigma_1, \sigma_2, \ldots, \sigma_n$ от $\theta_1, \ldots, \theta_n$:

$$\begin{aligned}
&\bar{\sigma}_1 = \sigma_1 - \theta_1,\\
&\bar{\sigma}_2 = \sigma_2 - \theta_1\bar{\sigma}_1 = \sigma_2 - \theta_1(\sigma_1 - \theta_1) = \sigma_2 - \sigma_1\theta_1 + \theta_1^2,\\
&\bar{\sigma}_3 = \sigma_3 - \theta_1\bar{\sigma}_2 = \sigma_3 - \theta_1(\sigma_2 - \sigma_1\theta_1 + \theta_1^2) = \sigma_3 - \sigma_2\theta_1 + \sigma_1\theta_1^2 - \theta_1^3
\end{aligned}$$

и т. д. Но по формулам Вьета

$$\sigma_1 = -a_1, \quad \sigma_2 = a_2, \ldots, \quad \sigma_n = (-1)^n a_n.$$

Следовательно,

$$\bar{\sigma}_1 = -a_1 - \theta_1, \quad \bar{\sigma}_2 = a_2 + a_1\theta_1 + \theta_1^2, \quad \bar{\sigma}_3 = -a_3 - a_2\theta_1 - a_1\theta_1^2 - \theta_1^3$$

и т. д.

Итак, мы видим, что $g(\theta_2) \ldots g(\theta_n)$ можно выразить через θ_1 и коэффициенты $a_1, \ldots, a_n$ многочлена $\varphi(x)$, т. е.

$$g(\theta_2)\,g(\theta_3) \ldots g(\theta_n) = k(\theta_1),$$

где $k(\theta_1)$ — многочлен от θ_1 над P. Отсюда

$$\frac{f(\theta_1)}{g(\theta_1)} = \frac{1}{b} f(\theta_1)\,k(\theta_1);$$

мы освободились от иррациональности в знаменателе.

Пример 1. Дана дробь

$$\frac{1}{1+\theta},$$

где θ — корень уравнения $x^3 - 2x - 2 = 0$. Освободиться от иррациональности в знаменателе.

Умножаем числитель и знаменатель дроби на $(1+\theta_2)(1+\theta_3)$, понимая под θ_2 и θ_3 остальные два корня рассматриваемого уравнения:

$$\frac{1}{1+\theta} = \frac{(1+\theta_2)(1+\theta_3)}{(1+\theta_1)(1+\theta_2)(1+\theta_3)},$$

где $\theta_1 = \theta$.

Обращаемся к симметрическому многочлену

$$F(\theta_1, \theta_2, \theta_3) = (1+\theta_1)(1+\theta_2)(1+\theta_3) = \\ = \theta_1\theta_2\theta_3 + (\theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3) + (\theta_1 + \theta_2 + \theta_3) + 1$$

и выражаем его через основные симметрические многочлены. Имеем:

$$\theta_1\theta_2\theta_3 = \sigma_3, \quad \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3 = \sigma_2, \quad \theta_1 + \theta_2 + \theta_3 = \sigma_1.$$

Отсюда

$$F(\theta_1, \theta_2, \theta_3) = \sigma_3 + \sigma_2 + \sigma_1 + 1.$$

Но по формулам Вьета имеем в данном случае, что $\sigma_1 = 0$, $\sigma_2 = -2$, $\sigma_3 = -(-2) = 2$. Следовательно,

$$F(\theta_1, \theta_2, \theta_3) = 2 - 2 + 0 + 1 = 1.$$

Таким образом,

$$\frac{1}{1+\theta} = (1+\theta_2)(1+\theta_3) = 1 + (\theta_2 + \theta_3) + \theta_2\theta_3.$$

Остаётся всё выразить через $\theta = \theta_1$. Пишем:

$$\theta_2 + \theta_3 = \sigma_1 - \theta_1,$$

$$\theta_2\theta_3 = \sigma_2 - \theta_1\theta_2 - \theta_1\theta_3 = \sigma_2 - \theta_1(\theta_2 + \theta_3) = \\ = \sigma_2 - \theta_1(\sigma_1 - \theta_1) = \sigma_2 - \sigma_1\theta_1 + \theta_1^2$$

или, так как $\sigma_1 = 0$, $\sigma_2 = -2$, $\sigma_3 = 2$:

$$\theta_2 + \theta_3 = -\theta_1, \quad \theta_2\theta_3 = -2 + \theta_1^2.$$

Отсюда окончательно получаем, что

$$\frac{1}{1+\theta} = 1 - \theta_1 - 2 + \theta_1^2 = -1 - \theta_1 + \theta_1^2 = -1 - \theta + \theta^2.$$

Пример 2. Избавиться от иррациональности в знаменателе дроби

$$\frac{2\sqrt[3]{5} - 1}{\sqrt[3]{25} + 4} \text{ *)}.$$

Здесь $\theta_1 = \sqrt[3]{5}$ есть корень уравнения $x^3 - 5 = 0$. Переписываем данную дробь так:

$$\frac{2\theta_1 - 1}{\theta_1^2 + 4}.$$

Умножаем, далее, числитель и знаменатель на $(\theta_2^2 + 4)(\theta_3^2 + 4)$:

$$\frac{2\theta_1 - 1}{\theta_1^2 + 4} = \frac{(2\theta_1 - 1)(\theta_2^2 + 4)(\theta_3^2 + 4)}{(\theta_1^2 + 4)(\theta_2^2 + 4)(\theta_3^2 + 4)}.$$

---

*) Этот пример легко решить, умножив числитель и знаменатель на $\sqrt[3]{25^2} - 4\sqrt[3]{25} + 16$; мы хотим, однако, продемонстрировать общий приём, указанный выше.

Симметрический многочлен

$$F(\theta_1, \theta_2, \theta_3) = (\theta_1^2+4)(\theta_2^2+4)(\theta_3^2+4) =$$
$$= \theta_1^2\theta_2^2\theta_3^2 + 4(\theta_1^2\theta_2^2 + \theta_1^2\theta_3^2 + \theta_2^2\theta_3^2) + 16(\theta_1^2+\theta_2^2+\theta_3^2) + 64$$

следующим образом выражается через основные симметрические многочлены:

$$F(\theta_1, \theta_2, \theta_3) = \sigma_3^2 + 4(\sigma_2^2 - 2\sigma_1\sigma_3) + 16(\sigma_1^2 - 2\sigma_2) + 64.$$

Так как в данном случае $\sigma_1 = 0$, $\sigma_2 = 0$, $\sigma_3 = 5$, то

$$F(\theta_1, \theta_2, \theta_3) = 25 + 64 = 89$$

и потому

$$\frac{2\theta_1 - 1}{\theta_1^2+4} = \frac{1}{89}(2\theta_1 - 1)(\theta_2^2+4)(\theta_3^2+4).$$

Выражаем

$$(\theta_2^2+4)(\theta_3^2+4) = \theta_2^2\theta_3^2 + 4(\theta_2^2+\theta_3^2) + 16$$

через $\bar{\sigma}_1$, $\bar{\sigma}_2$. Так как

$$\theta_2^2\theta_3^2 = (\theta_2\theta_3)^2 = \bar{\sigma}_2^2, \quad \theta_2^2+\theta_3^2 = (\theta_2+\theta_3)^2 - 2\theta_2\theta_3 = \bar{\sigma}_1^2 - 2\bar{\sigma}_2,$$

то

$$(\theta_2^2+4)(\theta_3^2+4) = \bar{\sigma}_2^2 + 4(\bar{\sigma}_1^2 - 2\bar{\sigma}_2) + 16.$$

Наконец, принимая во внимание, что

$$\bar{\sigma}_1 = -a_1 - \theta_1 = -\theta_1, \quad \bar{\sigma}_2 = a_2 + a_1\theta_1 + \theta_1^2 = \theta_1^2,$$

находим, что

$$(\theta_2^2+4)(\theta_3^2+4) = \theta_1^4 + 4(\theta_1^2 - 2\theta_1^2) + 16 =$$
$$= \theta_1^4 - 4\theta_1^2 + 16 = 5\theta_1 - 4\theta_1^2 + 16,$$

так как $\theta_1^3 = 5$. Итак,

$$\frac{2\theta_1 - 1}{\theta_1^2+4} = \frac{1}{89}(2\theta_1 - 1)(5\theta_1 - 4\theta_1^2 + 16) =$$
$$= \frac{1}{89}\cdot(-8\theta_1^3 + 14\theta_1^2 + 27\theta_1 - 16) = \frac{-40 + 14\theta_1^2 + 27\theta_1 - 16}{89} =$$
$$= \frac{14\theta_1^2 + 27\theta_1 - 56}{89}$$

или

$$\frac{2\sqrt[3]{5} - 1}{\sqrt[3]{25} + 4} = \frac{1}{89}(14\sqrt[3]{25} + 27\sqrt[3]{5} - 56).$$

2) Второе решение задачи об уничтожении иррациональности в знаменателе основано на использовании алгорифма Евклида и заключается в следующем.

Так как θ_1 — корень многочлена $\varphi(x)$ над полем P, то θ является числом, алгебраическим относительно P. Но ещё в § 10 мы

установили, что в случае такого θ_1 дробь

$$\frac{f(\theta_1)}{g(\theta_1)},$$

где $f(\theta_1)$ и $g(\theta_1) \neq 0$ — произвольные многочлены от θ_1 над P, может быть приведена к целому рациональному выражению от θ_1 с коэффициентами из P:

$$\frac{f(\theta_1)}{g(\theta_1)} = h(\theta_1),$$

где $h(\theta_1)$ — некоторый многочлен от θ_1 над P. В том же § 10 был указан метод нахождения многочлена $h(\theta_1)$. Отметим, что в этом методе существенно, чтобы многочлен $\varphi(x)$ был неприводим в поле P.

Пример 3. Освободиться от иррациональности в знаменателе дроби

$$\frac{\theta_1}{\theta_1 + 1},$$

где θ_1 — корень уравнения $\varphi(x) = x^5 - 2x - 2 = 0$. Здесь $g(x) = x + 1$ и многочлен $\varphi(x)$ неприводим в поле рациональных чисел.

Многочлен $\varphi(x) = x^5 - 2x - 2$ при делении на $g(x) = x + 1$ в частном даёт $q(x) = x^4 - x^3 + x^2 - x - 1$, а в остатке $r(x) = -1$. Таким образом,

$$\varphi(x) = g(x)\,q(x) - 1.$$

Отсюда

$$g(x)\,q(x) - \varphi(x) = 1.$$

Полагаем в последнем равенстве $x = \theta_1$:

$$g(\theta_1)\,q(\theta_1) = 1.$$

Следовательно, $q(\theta_1) = \frac{1}{g(\theta_1)}$ и потому

$$\frac{\theta_1}{1+\theta_1} = \theta_1 q(\theta_1) = \theta_1(\theta_1^4 - \theta_1^3 + \theta_1^2 - \theta_1 - 1) = \theta_1^5 - \theta_1^4 + \theta_1^3 - \theta_1^2 - \theta_1.$$

Так как θ_1 есть корень уравнения $x^5 - 2x - 2 = 0$, то $\theta_1^5 = 2\theta_1 + 2$, в силу чего

$$\frac{\theta_1}{1+\theta_1} = 2\theta_1 + 2 - \theta_1^4 + \theta_1^3 - \theta_1^2 - \theta_1 = -\theta_1^4 + \theta_1^3 - \theta_1^2 + \theta_1 + 2.$$

Симметрические многочлены находят также применение и в решении алгебраических уравнений.

Пусть

$$\varphi(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_n = 0 \qquad (3)$$

— уравнение n-й степени с комплексными коэффициентами и $\theta_1, \ldots, \theta_n$ — корни этого уравнения. Рассмотрим некоторое целое рациональное выражение

$$u = f(\theta_1, \theta_2, \ldots, \theta_n)$$

над полем рациональных чисел от корней $\theta_1, \theta_2, \ldots, \theta_n$. Будем про-

изводить всевозможные перестановки корней $\theta_1, \theta_2, \ldots, \theta_n$; при некоторых перестановках $\theta_1, \theta_2, \ldots, \theta_n$ u может не измениться, а при других перестановках $\theta_1, \ldots, \theta_n$ u может измениться. Пусть при этих перестановках u принимает m различных значений: $u_1 = u$, $u_2, \ldots, u_m$; очевидно, что $1 \leqslant m \leqslant n$. Составим следующий многочлен:

$$g(x) = (x - u_1)(x - u_2)\ldots(x - u_m) = \\ = x^m - g_1 x^{m-1} + g_2 x^{m-2} - \ldots + (-1)^m g_m,$$

где

$$g_1 = u_1 + u_2 + \ldots + u_m,$$
$$g_2 = u_1 u_2 + \ldots + u_{m-1} u_m,$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$g_m = u_1 u_2 \ldots u_m.$$

От перестановки $\theta_1, \theta_2, \ldots, \theta_n$ многочлен $g(x)$ не может измениться, может произойти лишь перестановка линейных множителей $x - u_i$. Отсюда ясно, что коэффициенты g_i многочлена $g(x)$ являются симметрическими многочленами корней $\theta_1, \ldots, \theta_n$. В силу этого коэффициенты g_i будут выражаться через коэффициенты данного уравнения (3):

$$g_i = h_i(a_1, \ldots, a_n),$$

где $h_i(a_1, \ldots, a_n)$ — многочлен от $a_1, \ldots, a_n$ над полем рациональных чисел. Уравнение

$$g(x) = (x - u_1)(x - u_2)\ldots(x - u_m) = \\ = x^m - g_1 x^{m-1} + g_2 x^{m-2} - \ldots + (-1)^m g_m = 0$$

называется *резольвентою* или *разрешающим* для уравнения (3).

В некоторых случаях удаётся с помощью соответствующего разрешающего уравнения свести решение данного уравнения (3) к решению уравнений более низких степеней. В качестве иллюстрации рассмотрим следующий пример:

Пример 4. Обратимся к уравнению четвёртой степени

$$x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0 \qquad (4)$$

и обозначим его корни через $\theta_1, \theta_2, \theta_3, \theta_4$. В качестве u мы возьмём

$$u = \theta_1\theta_2 + \theta_3\theta_4.$$

Легко видеть, что u при всевозможных перестановках корней $\theta_1, \theta_2, \theta_3, \theta_4$ принимает только три различных значения:

$$u_1 = u = \theta_1\theta_2 + \theta_3\theta_4, \quad u_2 = \theta_1\theta_3 + \theta_2\theta_4, \quad u_3 = \theta_1\theta_4 + \theta_2\theta_3.$$

Таким образом, разрешающее уравнение $g(x) = 0$ будет третьей степени. Найдём его коэффициенты g_1, g_2, g_3. С помощью сравни-

тельно несложных вычислений находим:

$$g_1 = u_1 + u_2 + u_3 = a_2,$$
$$g_2 = u_1u_2 + u_1u_3 + u_2u_3 = a_1a_3 - 4a_4,$$
$$g_3 = u_1u_2u_3 = a_3^2 + a_1^2a_4 - 4a_2a_4.$$

Следовательно, разрешающим уравнением будет:

$$x^3 - a_2x^2 + (a_1a_3 - 4a_4)x - (a_3^2 + a_1^2a_4 - 4a_2a_4) = 0\ *).$$

Теперь стоит только найти корни u_1, u_2, u_3 разрешающего уравнения, и мы легко определим корни данного уравнения (4). В самом деле, для u_1 имеем:

$$\theta_1\theta_2 + \theta_3\theta_4 = u_1, \quad \theta_1\theta_2 \cdot \theta_3\theta_4 = a_4,$$

откуда видно, что $\theta_1\theta_2$ и $\theta_3\theta_4$ являются корнями квадратного уравнения

$$x^2 - u_1x + a_4 = 0. \tag{5}$$

Далее, на основании формул Вьета

$$\theta_1\theta_2\theta_3 + \theta_1\theta_2\theta_4 + \theta_1\theta_3\theta_4 + \theta_2\theta_3\theta_4 = -a_3$$

или

$$\theta_1\theta_2(\theta_3 + \theta_4) + \theta_3\theta_4(\theta_1 + \theta_2) = -a_3.$$

Отсюда, обозначая корни $\theta_1\theta_2$ и $\theta_3\theta_4$ уравнения (5) соответственно через α и β, получаем следующие два соотношения для $\theta_1 + \theta_2$ и $\theta_3 + \theta_4$:

$$(\theta_1 + \theta_2) + (\theta_3 + \theta_4) = -a_1, \quad \beta(\theta_1 + \theta_2) + \alpha(\theta_3 + \theta_4) = -a_3.$$

Из этих соотношений без труда находим, что

$$\theta_1 + \theta_2 = \frac{a_3 - \alpha a_1}{\alpha - \beta}, \quad \theta_3 + \theta_4 = \frac{\beta a_1 - a_3}{\alpha - \beta}.$$

Следовательно, θ_1 и θ_2 являются корнями квадратного уравнения

$$x^2 - \frac{a_3 - \alpha a_1}{\alpha - \beta}x + \alpha = 0,$$

а θ_3 и θ_4 — корнями квадратного уравнения

$$x^2 - \frac{\beta a_1 - a_3}{\alpha - \beta}x + \beta = 0.$$

---

*) Это уравнение совпадает с разрешающим уравнением (3), построенным по способу Феррари (стр. 221), если положить $x = 2y$.

ГЛАВА III

О РЕШЕНИИ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ В РАДИКАЛАХ

§ 15. Подстановки

В настоящей главе мы покажем, что алгебраические уравнения выше четвёртой степени вообще нельзя решить в радикалах. Однако для этой цели нам придётся предварительно познакомиться с понятием подстановки, с понятием, представляющим также и самостоятельный интерес.

Пусть

$$a_1,\ a_2,\ \ldots,\ a_n \tag{1}$$

— некоторая совокупность n элементов. Что собой представляют эти элементы, нас не интересует.

Подстановкой n-й степени элементов (1) называется такая замена каждого элемента a_i некоторым элементом a_j той же совокупности (1), при которой различные элементы переходят также в различные элементы. (В частности, тот или иной элемент a_i может быть заменён самим собою, т. е. может быть оставлен без изменения.)

Обычно подстановку n-й степени записывают следующим образом:

$$S=\begin{pmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_{i_1} & a_{i_2} & a_{i_3} & \ldots & a_{i_n} \end{pmatrix}.$$

Здесь под элементами $a_1,\ a_2,\ \ldots,\ a_n$ первой строки находятся соответственно элементы $a_{i_1},\ a_{i_2},\ \ldots,\ a_{i_n}$ второй строки. Это означает, что a_1 заменяется элементом a_{i_1}, a_2 — элементом a_{i_2}, $\ldots$, a_n — элементом a_{i_n}.

Например,

$$S=\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_3 & a_4 & a_1 & a_2 \end{pmatrix} \tag{2}$$

есть подстановка четвёртой степени элементов $a_1,\ a_2,\ a_3,\ a_4$. Мы здесь видим, что под элементом a_1 первой строки находится элемент a_3 второй строки. Это значит, что подстановка S заменяет

элемент a_1 элементом a_3. Далее, из записи подстановки (2) видно, что эта подстановка заменяет элемент a_2 элементом a_4, a_3 — элементом a_1 и a_4 — элементом a_2.

Условимся для упрощения писать вместо элементов $a_1, a_2, \ldots, a_n$ их номера и говорить о подстановках n-й степени из n чисел $1, 2, \ldots, n$. Так, например, вместо подстановки (2) можно говорить о подстановке

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}.$$

Введём теперь операцию умножения подстановок n-й степени. Для большей наглядности возьмём две подстановки четвёртой степени:

$$S=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \quad \text{и} \quad T=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}. \tag{3}$$

Посмотрим, что получится, если сначала применить подстановку S и затем подстановку T.

Подстановка S число 1 переводит в число 3, после чего 3 переводится подстановкой T в 4. Таким образом, при последовательном применении подстановок S и T число 1 переводится в 4:

$$1 \to 4.$$

Далее, число 2 подстановкой S переводится в 1, после чего 1 подстановкой T переводится в 2. Таким образом, при последовательном применении подстановок S и T число 2 переводится в 2:

$$2 \to 2,$$

т. е. остаётся без изменения. Подобным же образом находим, что при последовательном применении подстановок S и T число 3 переводится в 1 и число 4 переводится в 3:

$$3 \to 1,$$

$$4 \to 3.$$

Мы видим отсюда, что последовательное применение подстановок S и T равносильно применению одной подстановки

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}. \tag{4}$$

Подстановку (4) мы назовём *произведением* подстановок S и T и обозначим через ST.

Вообще произведением S_1S_2 подстановок S_1 и S_2 n-й степени называется такая подстановка n-й степени, которая равносильна последовательному применению подстановок S_1 и S_2.

Легко убедиться, что умножение подстановок не подчиняется переместительному (коммутативному) закону. Например, произведение TS подстановок (3) будет уже отлично от подстановки (4)

$$TS=\begin{pmatrix}1 & 2 & 3 & 4\\ 1 & 4 & 2 & 3\end{pmatrix}.$$

Относительно умножения подстановок справедлива следующая

Теорема 26. *Множество S_n всех подстановок n-й степени образует группу относительно операции умножения подстановок.*

Доказательство. Прежде всего покажем, что умножение подстановок подчиняется сочетательному закону: для любых трёх подстановок S_1, S_2 и S_3 n-й степени имеет место равенство

$$(S_1S_2)\,S_3=S_1\,(S_2S_3).$$

В самом деле, пусть α — одно из чисел 1, 2, ... , n и пусть подстановка S_1 число α переводит в β, подстановка S_2 число β переводит в γ и γ подстановкою S_3 переводится в δ. Тогда $(S_1S_2)\,S_3$ будет α переводить в δ, так как S_1S_2 число α переводит в γ, после чего S_3 число γ переводит в δ.

Точно так же $S_1\,(S_2S_3)$ переводит α в δ, так как S_1 переводит α в β и S_2S_3 переводит β в δ. Мы видим отсюда, что $(S_1S_2)\,S_3$ и $S_1\,(S_2S_3)$ производят одинаковую замену — они α переводят в δ. Следовательно,

$$(S_1S_2)\,S_3=S_1\,(S_2S_3).$$

Теперь покажем, что среди подстановок n-й степени имеется подстановка, играющая роль правого единичного элемента. Для этой цели рассмотрим подстановку

$$I=\begin{pmatrix}1 & 2 & 3 & \dots & n\\ 1 & 2 & 3 & \dots & n\end{pmatrix},$$

носящую название *тождественной* или *единичной подстановки.* Нетрудно проверить, что для любой подстановки n-й степени

$$S=\begin{pmatrix}1 & 2 & 3 & \dots & n\\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_n\end{pmatrix} \tag{5}$$

выполняется равенство

$$SI=S.$$

Следовательно, подстановка I и будет правой единицей.

Наконец, покажем, что для каждой подстановки (5) существует подстановка той же степени, играющая для S роль правого обратного элемента. Мы утверждаем, что такой подстановкой будет:

$$S'=\begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n\\ 1 & 2 & \dots & n\end{pmatrix}.$$

Действительно, 1 переводится подстановкою S в α_1, после чего S' переводит α_1 в 1; следовательно, произведение SS' переводит 1 в 1. Подобным же образом убеждаемся, что произведение SS' переводит 2 в 2, 3 в 3, ..., n в n. Значит, $SS'=I$, что и требовалось показать.

Подстановка S' обозначается через S^{-1} и называется *обратною* относительно S.

Группа S_n всех подстановок n-й степени обычно называется *симметрическою группой* n-й степени; она является некоммутативной группой, так как вообще умножение подстановок не обладает свойством переместительности. Симметрическая группа S_n состоит, очевидно, из $n!$ подстановок, вследствие чего S_n есть конечная группа порядка, равного $n!$.

Всякая подгруппа G симметрической группы S_n называется *группой подстановок* n-й степени. Например, существует всего шесть групп подстановок третьей степени, а именно сама симметрическая группа S_3 третьей степени; группа G_1, состоящая из трёх подстановок:

$$I=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\quad \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix},\quad \begin{pmatrix}1&2&3\\3&1&2\end{pmatrix};$$

группа G_2, состоящая из двух подстановок:

$$I=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\quad \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix};$$

группа G_3, состоящая из двух подстановок:

$$I=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\quad \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix};$$

группа G_4, состоящая из двух подстановок:

$$I=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix},\quad \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix};$$

наконец, группа G_5, состоящая лишь из единичной подстановки

$$I=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}.$$

Группу G_5 принято обозначать через E и называть единичной группой.

§ 16. О неразрешимости уравнений выше четвёртой степени в радикалах

В настоящем параграфе мы увидим, что для всех алгебраических уравнений данной степени $n \geqslant 5$ не существует общей формулы решения в радикалах.

В § 10 было показано, что уравнение

$$a_0 x^n + a_1 x^{n-1} + \ldots + a_n = 0 \quad (a_0 \neq 0) \tag{1}$$

тогда и только тогда разрешимо в радикалах, когда нормальное поле $\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$ содержится в расширении $\Delta(\rho_1, \rho_2, \ldots, \rho_k)$, полученном путём присоединения к Δ некоторых радикалов $\rho_1 = \sqrt[n_1]{A_1}$, $\rho_2 = \sqrt[n_2]{A_2}$, $\ldots$, $\rho_k = \sqrt[n_k]{A_k}$, где A_1 принадлежит Δ, A_2 принадлежит $\Delta(\rho_1)$, $\ldots$, A_k принадлежит $\Delta(\rho_1, \rho_2, \ldots, \rho_{k-1})$; при этом поле Δ есть не что иное, как область рациональности уравнения (см. § 10).

Мы можем всегда показатели $n_1, n_2, \ldots, n_k$ радикалов предполагать простыми числами. В самом деле, если бы встретился, например, радикал $\sqrt[12]{A}$, то мы его заменили бы тремя радикалами $\rho' = \sqrt{A}$, $\rho'' = \sqrt{\rho'}$ и $\rho''' = \sqrt[3]{\rho''}$. В знак этого мы будем в дальнейшем показатель радикала ρ_i обозначать не через n_i, а через p_i и будем молчаливо под p_i подразумевать простое число.

Присоединим к Δ первообразные корни p_1-й, p_2-й, $\ldots$, p_k-й степени из единицы и обозначим через K поле, получившееся в результате такого присоединения: $K = \Delta(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$. Очевидно, что если уравнение (1) разрешимо в радикалах, то его нормальное поле Ω будет и подавно содержаться в расширении $K(\rho_1, \rho_2, \ldots, \rho_k)$. Однако часть радикалов $\rho_1 = \sqrt[p_1]{A_1}$, $\ldots$, $\rho_k = \sqrt[p_k]{A_k}$ может оказаться лишней. А именно, если подкоренная величина A_i радикала $\rho_i = \sqrt[p_i]{A_i}$ будет точной p_i-й степенью в поле $K(\rho_1, \ldots, \rho_{i-1})$, т. е. если $A_i = a^{p_i}$, где a — некоторый элемент из $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$, то в этом поле будут лежать все корни двучлена $x^{p_i} - A_i$, в силу чего радикал ρ_i будет лишним — его присоединение к полю $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$ никакого расширения фактически не даст:

$$K(\rho_1, \rho_2, \ldots, \rho_i) = K(\rho_1, \rho_2, \ldots, \rho_{i-1}).$$

Теперь докажем следующие леммы:

Лемма 1. *Если A_i не является точной p_i-й степенью в поле $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$, то двучлен $x^{p_i} - A_i$ неприводим в поле $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$.*

Доказательство. Предположим противное, пусть этот двучлен приводим в поле $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$:

$$x^{p_i} - A_i = \varphi(x)\,\psi(x),$$

где $\varphi(x)$ и $\psi(x)$ — многочлены над полем $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$. Обозначим через ε первообразный корень p_i-й степени из единицы и через θ_0 — какой-нибудь корень нашего двучлена. Тогда, как известно, любой корень θ_ν двучлена найдётся по формуле

$$\theta_\nu = \varepsilon^\nu \theta_0.$$

Отсюда свободный член b многочлена $\varphi(x)$ будет равен

$$b = (-1)^r \theta_{\nu_1} \theta_{\nu_2} \ldots \theta_{\nu_r} = \varepsilon' (-\theta_0)^r,$$

где $\varepsilon' = \varepsilon^{\nu_1 + \cdots + \nu_r}$ и $1 \leqslant r < p_i$. Очевидно, что ε' есть некоторый корень p_i-й степени из единицы. Возведём b в p_i-ю степень:

$$b^{p_i} = \varepsilon'^{p_i} (-\theta_0)^{rp_i} = (-1)^{rp_i} A_i^r, \text{ т. е. } A_i^r = (-1)^{rp_i} b^{p_i}.$$

Так как $1 \leqslant r < p_i$ и p_i — простое число, то r и p_i взаимно просты; отсюда существуют такие целые числа s и t, что $rs + p_i t = 1$. Таким образом, получаем:

$$A_i = A_i^{rs + p_i t} = A_i^{rs} A_i^{p_i t} = (-1)^{rp_i s} b^{p_i s} A_i^{p_i t} = [(-1)^{rs} b^s A_i^t]^{p_i},$$

т. е. A_i является точной p_i-й степенью в поле $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$, что невозможно.

Лемма 2. *Если радикал* $\rho_i = \sqrt[p_i]{A_i}$ *не лежит в поле* $K(\rho_1, \ldots, \rho_{i-1})$, *то целая степень* ρ_i^m *тогда и только тогда лежит в поле* $K(\rho_1, \ldots, \rho_{i-1})$, *когда* m *делится на* p_i.

Доказательство. Если m делится на p_i, то $m = p_i q$, где q — некоторое целое число. Отсюда

$$\rho_i^m = \rho_i^{p_i q} = (\rho_i^{p_i})^q = A_i^q.$$

Но A_i и тем самым A_i^q лежат в поле $K(\rho_1, \ldots, \rho_{i-1})$. Следовательно, ρ_i^m должно лежать в $K(\rho_1, \ldots, \rho_{i-1})$.

Обратно, пусть ρ_i^m лежит в $K(\rho_1, \ldots, \rho_{i-1})$, т. е. $\rho_i^m = a$, где a — некоторый элемент $K(\rho_1, \ldots, \rho_{i-1})$. Обозначим через q частное и через r остаток при делении m на p_i. Тогда мы можем написать, что $m = p_i q + r$. Предположим, что остаток r отличен от нуля. В таком случае

$$\rho_i^m = \rho_i^{p_i q + r} = (\rho_i^{p_i})^q \rho_i^r = A_i^q \rho_i^r,$$

откуда в силу равенства $\rho_i^m = a$ получаем, что

$$A_i^q \rho_i^r = a \quad \text{или} \quad \rho_i^r = b,$$

где $b = a A_i^{-q}$ — элемент поля $K(\rho_1, \ldots, \rho_{i-1})$. Мы видим отсюда, что ρ_i является одновременно корнем многочлена $p(x) = x^{p_i} - A_i$ и многочлена $\varphi(x) = x^r - b$, в силу чего многочлены $p(x)$ и $\varphi(x)$ не взаимно просты. Но по лемме 1 многочлен $p(x)$ неприводим

в поле $K(\rho_1, \ldots, \rho_{i-1})$. Следовательно, поскольку многочлены $p(x)$ и $\varphi(x)$ не взаимно просты, $\varphi(x)$ должно делиться на $p(x)$. Но это невозможно, так как степень r многочлена $\varphi(x)$ меньше степени p_i многочлена $p(x)$. Поэтому предположение, что $r \neq 0$ неверно, т. е. m делится на p_i.

Лемма 3. *Если уравнение* (1) *разрешимо в радикалах, то каждый из корней уравнения можно следующим образом выразить через радикалы*:

$$\alpha = u_0 + \rho_h + u_2\rho_h^2 + \ldots + u_{p_h-1}\rho_h^{p_h-1},$$

где $\rho_1 = \sqrt[p_1]{A_1}$, $\rho_2 = \sqrt[p_2]{A_2}$, $\ldots$, $\rho_h = \sqrt[p_h]{A_h}$ (p_i — *простые числа*), A_1 — *элемент поля* K; A_2 — *элемент поля* $K(\rho_1)$, $\ldots$, A_h — *элемент поля* $K(\rho_1, \ldots, \rho_{h-1})$; K — *расширение, получающееся путём присоединения к области рациональности* Δ *уравнения первообразных корней* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ *соответственно* p_1-й, $\ldots$, p_k-й *степени из единицы* ($k \geqslant h$); u_i — *элементы* $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$. *При этом* ρ_1 *не лежит в* K, $\ldots$, ρ_h *не лежит в* $K(\rho_1, \ldots, \rho_{h-1})$ *и* α *не лежит в* $K(\rho_2, \rho_3, \ldots, \rho_h)$, $K(\rho_1, \rho_3, \ldots, \rho_h)$, $\ldots$, $K(\rho_1, \rho_2, \ldots, \rho_{h-2}, \rho_h)$, $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$.

Доказательство. Пусть $\alpha = \alpha_1$ — один из корней уравнения (1). Так как уравнение (1) разрешимо в радикалах, то α будет содержаться в $\Omega = \Delta(\rho_1, \ldots, \rho_k)$ и, следовательно, будет содержаться в одном из полей вида $K(\rho_1, \rho_2, \ldots, \rho_h)$ ($h \leqslant k$). Отсюда следует, что

$$\alpha = a_0 + a_1\rho_h + a_2\rho_h^2 + \ldots + a_{p_h-1}\rho_h^{p_h-1}, \tag{2}$$

где a_i — элементы поля $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$. Мы можем предположить, что ρ_1 не лежит в K, ρ_2 не лежит в $K(\rho_1)$, $\ldots$, ρ_h не лежит в $K(\rho_1, \ldots, \rho_{h-1})$ и α не лежит в $K(\rho_1, \ldots, \rho_{i-1}, \rho_{i+1}, \ldots, \rho_h)$ ($i = 1, 2, \ldots, h$). В самом деле, если бы имело место иное, то в выражении (2) некоторые радикалы ρ_i можно было бы опустить.

Покажем, что при соответствующем выборе радикала ρ_h величина a_1 может быть сделана равной единице. Действительно, все $a_1, \ldots, a_{p_h-1}$ в правой части равенства (2) не могут равняться нулю, так как в противном случае α лежало бы в $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$, что невозможно. Поэтому пусть $a_l \neq 0$ ($1 \leqslant l < p_h$). Тогда полагаем:

$$a_l\rho_h^l = \rho_h'.$$

Так как числа l и p_h взаимно просты, то существуют такие целые числа s и t, что $sl + tp_h = 1$. Легко видеть, что s не делится на p_h: если бы s делилось на p_h, то, очевидно, и $sl + tp_h$ делилось бы на p_h, вследствие чего и 1 делилась бы на простое число p_h, а это невозможно. Возведём ρ_h' в s-ю степень:

$$\rho_h'^s = a_l^s\rho_h^{ls} = a_l^s\rho_h^{1-tp_h} = a_l^s\rho_h A_h^{-t},$$

откуда

$$\rho_h = v\rho_h'^s,$$

где $v = A_h^t a_l^{-s}$ — элемент, лежащий в $K(\rho_1, \ldots, \rho_{h-1})$. Очевидно, что ρ'_h не лежит в $K(\rho_1, \ldots, \rho_{h-1})$; в противном случае $\rho_h = v\rho_h'^s$ и подавно лежало бы в $K(\rho_1, \ldots, \rho_{h-1})$. Таким образом, мы можем вместо ρ_h взять в качестве h-го радикала ρ'_h. Заменяя в равенстве (2) радикал ρ_h его выражением через ρ'_h и принимая во внимание, что $a_l \rho_h'^l = \rho'_h$, получаем:

$$\alpha = \alpha_1 = a_0 + a_1 v \rho_h'^s + a_2 v^2 \rho_h'^{2s} + \ldots + \rho'_h + \ldots + a_{p_h-1} v^{p_h-1} \rho_h'^{(p_h-1)s}. \quad (3)$$

В равенстве (3) все степени $\rho_h'^{\nu s}$ ($\nu = 0, 1, \ldots, p_h - 1$) между собой различны. В самом деле, если бы

$$\rho_h'^{\nu_1 s} = \rho_h'^{\nu_2 s} \qquad (\nu_1 > \nu_2),$$

то $\rho_h'^{(\nu_1 - \nu_2)s} = 1$, откуда по лемме 2 $(\nu_1 - \nu_2)s$ делилось бы на p_h, т. е. $\nu_1 - \nu_2$ делилось бы на p_h, так как s не делится на простое число p_h. Но $\nu_1 - \nu_2$ не может делиться на p_h, так как $0 < \nu_1 - \nu_2 < p_h$. Следовательно, при $\nu_1 \neq \nu_2$ степени $\rho_h'^{\nu_1 s}$ и $\rho_h'^{\nu_2 s}$ различны.

Пусть, далее, q — частное и r — остаток при делении νs на p_h. Тогда

$$\rho_h'^{\nu s} = (\rho_h'^{p_h})^q \rho_h'^r = b\rho_h'^r,$$

где $b = (\rho'^{p_h})^q$ — элемент поля $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$. Очевидно, что при изменении ν от 1 до $p_h - 1$ остаток r будет в той или иной последовательности принимать $p_h - 1$ различных значений $1, 2, \ldots, p_h - 1$. Отсюда равенство (3) принимает следующий вид:

$$\alpha = u_0 + \rho'_h + u_2 \rho_h'^2 + \ldots + u_{p_h-1} \rho_h'^{p_h-1},$$

т. е. вместо a_1 получилась единица.

Лемма 4. *Если уравнение* (1) *разрешимо в радикалах* $\rho_1, \rho_2, \ldots, \rho_k$ *с показателями* $p_1, p_2, \ldots, p_k$, *то радикалы* ρ_i *являются целыми рациональными функциями корней уравнения* (1) *над полем* K, *где* K *имеет тот же смысл, что и в лемме* 3.

Доказательство. Пусть $\alpha = \alpha_1$ — какой-нибудь корень уравнения (1). Согласно лемме 3 мы можем написать, что

$$\alpha = \alpha_1 = u_0 + \rho_h + u_2 \rho_h^2 + \ldots + u_{p_h-1} \rho_h^{p_h-1}, \quad (4)$$

причём ρ_i не лежит в поле $K(\rho_1, \rho_2, \ldots, \rho_{i-1})$ и α не лежит в поле $K(\rho_1, \ldots, \rho_{i-1}, \rho_{i+1}, \ldots, \rho_h)$ $(i = 1, 2, \ldots, h)$. Так как $\rho_h = \sqrt[p_h]{A_h}$ не лежит в поле $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$, то по лемме 1 двучлен $x^{p_h} - A_h$ неприводим в поле $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$. Принимая это во внимание, подставим в уравнение (1) вместо x выражение (4) корня α. Тогда в левой части уравнения получится относительно ρ_h многочлен, степень которого с помощью равенства $\rho_h^{p_h} = A_h$ можно понизить до $p_h - 1$:

$$B_0 + B_1 \rho_h + \ldots + B_{p_h-1} \rho_h^{p_h-1} = 0,$$

где B_i лежат в $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$. В силу неприводимости двучлена $x^{p_h} - A_h$ в поле $K(\rho_1, \rho_2, \ldots, \rho_{h-1})$ последнее равенство может иметь место только тогда, когда $B_0 = B_1 = \ldots = B_{p_h-1} = 0$. Таким образом, это равенство будет осуществляться для любого корня $\varepsilon_h^{\mu}\rho_h$ двучлена $x^{p_h} - A_h$. Следовательно,

$$\alpha_{\mu+1} = u_0 + \varepsilon_h^{\mu}\rho_h + \ldots + u_{p_h-1}\,\varepsilon_h^{\mu(p_h-1)}\,\rho_h^{p_h-1} \tag{5}$$

$$(\mu = 0, 1, \ldots, p_h - 1)$$

есть также корень уравнения (5).

Умножим теперь каждое из равенств (5) на $\varepsilon_h^{-\mu\nu}$ и сложим почленно $(1 \leqslant \nu \leqslant p_h - 1)$. После некоторых упрощений у нас получится:

$$p_h\rho_h = \sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu}\alpha_{\mu+1}, \qquad p_h u_\nu \rho_h^\nu = \sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu\nu}\alpha_{\mu+1} \quad (\nu = 2, \ldots, p_h - 1),$$

откуда

$$\rho_h = \frac{1}{p_h}\sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu}\alpha_{\mu+1}, \qquad u_\nu = p_h^{\nu-1}\left(\sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu\nu}\alpha_{\mu+1}\right)\left(\sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu}\alpha_{\mu+1}\right)^{-\nu},$$

т. е. ρ_h и u_ν лежат в $K(\alpha_1, \ldots, \alpha_n)$.

Таким образом, величины $A_h = \rho_h^{p_h}$ и u_ν по только что доказанному являются целыми рациональными функциями корней $\alpha_1, \ldots, \alpha_n$ над полем K, а с другой стороны, выражаются через радикалы $\rho_1, \ldots, \rho_{h-1}$. Введём для этих величин с целью сокращения письма единое обозначение β_ν. По меньшей мере одна из величин β_ν должна содержать радикал ρ_{h-1}. Если бы это было не так, то в выражении (4) корня α через радикалы можно было бы ρ_{h-1} опустить, вследствие чего мы имели бы, что корень α лежит в $K(\rho_1, \ldots, \rho_{h-2}, \rho_h)$, что невозможно. Пусть β_1 содержит ρ_{h-1}. Напишем выражение величины β_1 через радикалы:

$$\beta_1 = v_0 + v_1\rho_{h-1} + v_2\rho_{h-1}^2 + \ldots + v_{p_{h-1}-1}\,\rho_{h-1}^{p_{h-1}-1}. \tag{6}$$

Так как β_1 есть целая рациональная функция корней $\alpha_1, \ldots, \alpha_n$ над K:

$$\beta_1 = r(\alpha_1, \ldots, \alpha_n),$$

то мы можем в выражении $r(\alpha_1, \ldots, \alpha_n)$ произвести всевозможные перестановки корней α_j, в результате чего получим $n!$ значений: $\theta_1 = \beta_1, \theta_2, \ldots, \theta_{n!}$. Составим уравнение

$$g(x) = \prod_{i=1}^{n!}(x - \theta_i) = 0. \tag{7}$$

Очевидно, что коэффициенты этого уравнения суть симметрические многочлены от $\alpha_1, \alpha_2, \ldots, \alpha_n$ над полем K; поэтому $g(x)$ есть многочлен над K.

Итак, мы видим, что β_1 есть корень уравнения (7), разрешимого в радикалах. Согласно лемме 3 в выражении (6) можно принять $v_1 = 1$:

$$\beta_1 = v_0 + \rho_{h-1} + \ldots + v_{p_{h-1}-1}\rho_{h-1}^{p_{h-1}-1}.$$

Повторяя относительно β_1 те же рассуждения, что и для α_1, получаем, что ρ_{h-1} и v_ν лежат в $K(\alpha_1, \ldots, \alpha_n)$.

Далее, вводим для величин ρ_{h-1} и v_ν единое обозначение γ_ν. По меньшей мере одна из величин γ_ν, например γ_1, должна содержать радикал ρ_{h-2}. Повторяя для γ_1 рассуждения, аналогичные тем, которые мы провели для β_1, получим, что ρ_{h-2} лежит в $K(\alpha_1, \ldots, \alpha_n)$ и т. д. В конечном счёте мы дойдём до радикала ρ_1 и покажем, что он лежит в $K(\alpha_1, \ldots, \alpha_n)$, и этим завершим доказательство леммы.

Лемма 5. *Пусть* $T = R(\varepsilon_1, \ldots, \varepsilon_k)$ *— расширение, получающееся путём присоединения к полю* R *рациональных чисел первообразных корней* $\varepsilon_1, \ldots, \varepsilon_k$, *соответственно* p_1*-й,* p_2*-й,* $\ldots$, p_k*-й степени из единицы. Тогда всякое рациональное соотношение над полем* T:

$$\varphi(x_1, x_2, \ldots, x_n, \sigma_1, \sigma_2, \ldots, \sigma_n) = 0 \qquad (8)$$

между независимыми переменными $x_1, \ldots, x_n$ *и основными симметрическими многочленами* $\sigma_1, \sigma_2, \ldots, \sigma_n$ *от этих переменных остаётся в силе при любой перестановке переменных*:

$$\varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_n}, \sigma_1, \sigma_2, \ldots, \sigma_n) = 0, \qquad (9)$$

где $i_1, i_2, \ldots, i_n$ *— произвольная перестановка чисел* $1, 2, \ldots, n$.

Доказательство. Для произвольной системы значений переменных $x_1 = \alpha_1, \ldots, x_n = \alpha_n$ (α_i — любые комплексные числа) пусть $\sigma_1 = p_1, \ldots, \sigma_n = p_n$. Тогда согласно условиям леммы

$$\varphi(\alpha_1, \alpha_2, \ldots, \alpha_n, p_1, p_2, \ldots, p_n) = 0.$$

Полагая, далее, $x_1 = \alpha_{i_1}, x_2 = \alpha_{i_2}, \ldots, x_n = \alpha_{i_n}$, мы при этих значениях переменных будем иметь, очевидно, те же значения $\sigma_1 = p_1, \ldots, \sigma_n = p_n$ основных симметрических многочленов. Таким образом, мы можем в соотношении (8) положить $x_1 = \alpha_{i_1}, x_2 = \alpha_{i_2}, \ldots, x_n = \alpha_{i_n}, \sigma_1 = p_1, \ldots, \sigma_n = p_n$, в результате чего получим:

$$\varphi(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}, p_1, p_2, \ldots, p_n) = 0.$$

Отсюда в силу произвольности чисел $\alpha_1, \alpha_2, \ldots, \alpha_n$ следует, что

$$\varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_n}, \sigma_1, \ldots, \sigma_n) = 0.$$

Замечание. Переход от соотношения (8) к соотношению (9) можно, очевидно, осуществить с помощью подстановки

$$\begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}.$$

Мы можем, следовательно, сказать, что *соотношение* (8) *не нарушается при любой подстановке симметрической группы* S_n *n-й степени.* Теперь мы вплотную подошли к известной теореме о неразрешимости алгебраических уравнений выше четвёртой степени в радикалах.

Теорема Руффини-Абеля. *Для алгебраических уравнений заданной n-й степени не существует общих формул, выражающих каждый корень уравнения через радикалы, когда* $n \geqslant 5$.

Доказательство. Предположим противное — пусть какой-нибудь корень x_1 произвольного алгебраического уравнения

$$x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \dots + (-1)^n \sigma_n = 0$$

степени $n \geqslant 5$ выражается через радикалы по общей формуле

$$x_1 = r(\rho_1, \dots, \rho_h, \sigma_1, \sigma_2, \dots, \sigma_n), \tag{10}$$

где $r(\rho_1, \dots, \rho_h, \sigma_1, \sigma_2, \dots, \sigma_n)$ — рациональная функция от $\rho_1, \dots, \rho_h, \sigma_1, \dots, \sigma_n$ над полем $T = R(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)$, не зависящая от выбора уравнения; ε_i, как и выше, означает первообразный корень p_i-й степени из единицы. В силу произвольности алгебраического уравнения мы можем его корни $x_1, \dots, x_n$ рассматривать как независимые переменные. Согласно лемме 4 радикалы ρ_i должны быть целыми рациональными функциями от $x_1, \dots, x_n$ над полем $K = \Delta(\varepsilon_1, \dots, \varepsilon_k)$*) или, что то же, рациональными функциями от $x_1, \dots, x_n, \sigma_1, \dots, \sigma_n$ над полем T:

$$\rho_i = r_i(x_1, \dots, x_n, \sigma_1, \dots, \sigma_n), \tag{11}$$

причём, поскольку формула (10) является общей для всех алгебраических уравнений заданной степени $n \geqslant 5$, выражения

$$r_i(x_1, \dots, x_n, \sigma_1, \dots, \sigma_n)$$

также не должны зависеть от выбора уравнения заданной степени n.

Рассмотрим подстановку

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \dots & n \\ 2 & 3 & 4 & 5 & 1 & 6 & \dots & n \end{pmatrix}.$$

Она переводит корень x_1 в x_2, x_2 в x_3, x_3 в x_4, x_4 в x_5 и x_5 в x_1, а остальные корни (при $n > 5$) оставляет неподвижными. Будем обозначать через Ht выражение, получающееся из некоторой рациональной функции H от $x_1, \dots, x_n, \sigma_1, \dots, \sigma_n$ над полем T

*) $\Delta = R(\sigma_1, \sigma_2, \dots, \sigma_n)$.

в результате применения подстановки t n-й степени. Покажем, что от подстановки s величина первого радикала не меняется.

Так как $\rho_1 = r_1(x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n) = \sqrt[p_1]{A_1}$ и A_1 есть рациональная функция над полем T от $\sigma_1, \sigma_2, \ldots, \sigma_n$, то равенство

$$\rho_1^{p_1} = A_1$$

можно рассматривать как рациональное соотношение между $x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n$ над полем T; таким образом, это соотношение по лемме 5 не нарушится после применения подстановки s:

$$(\rho_1^{p_1})\, s = A_1 s$$

или, так как $(\rho_1^{p_1})\, s = (\rho_1 s)^{p_1}$ и $A_1 s = A_1$, имеем:

$$(\rho_1 s)^{p_1} = A_1,$$

т. е. $\rho_1 s$ оказалось также корнем p_1-й степени из A_1. Отсюда $\rho_1 s = \varepsilon_1^{\nu}\rho_1$, где ν — некоторое целое неотрицательное число. Далее, $\rho_1 s^m = (\rho_1 s)\, s^{m-1} = \varepsilon_1^{\nu}(\rho_1 s^{m-1}) = \ldots = \varepsilon_1^{m\nu}\rho_1$. Но $s^5 = I$, где I — единичная подстановка. Следовательно, $\rho_1 s^5 = \rho_1 = \varepsilon_1^{5\nu}\rho_1$, откуда $\varepsilon_1^{5\nu} = 1$.

Теперь обратимся к подстановкам

$$t = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 1 & 2 & 4 & 5 & 3 & 6 & \ldots & n \end{pmatrix} \quad \text{и} \quad u = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 2 & 3 & 1 & 4 & 5 & 6 & \ldots & n \end{pmatrix}.$$

Легко видеть, что $tu = s$ и $t^3 = u^3 = I$. Рассуждая аналогичным образом, получаем, что $\rho_1 t = \varepsilon_1^{\mu}\rho_1$ и $\rho_1 u = \varepsilon_1^{\lambda}\rho_1$, где μ и λ — некоторые целые неотрицательные числа и $\varepsilon_1^{3\mu} = \varepsilon_1^{3\lambda} = 1$. Отсюда

$$\rho_1 s = \rho_1 (tu) = \varepsilon_1^{\mu}(\rho_1 u) = \varepsilon_1^{\mu} \cdot \varepsilon_1^{\lambda}\rho_1 = \varepsilon_1^{\mu + \lambda}\rho_1,$$

вследствие чего $\varepsilon_1^{\nu} = \varepsilon_1^{\mu + \lambda}$. А теперь получаем, что

$$\varepsilon_1^{\nu} = \frac{\varepsilon_1^{6\nu}}{\varepsilon_1^{5\nu}} = \frac{\varepsilon_1^{6\mu + 6\lambda}}{1} = \frac{(\varepsilon_1^{3\mu}\varepsilon_1^{3\lambda})^2}{1} = \frac{1}{1} = 1,$$

т. е. $\rho_1 s = \rho_1$ — подстановка s оставляет радикал ρ_1 без изменения.

Переходя к ρ_2, затем к ρ_3 и т. д., убеждаемся с помощью аналогичных рассуждений, что все радикалы $\rho_1, \rho_2, \ldots, \rho_h$ не меняются от подстановки s.

Наконец, обращаемся к равенству (10). Согласно равенствам (11) мы можем равенство (10) рассматривать как рациональное соотношение между $x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n$ над полем T. Таким образом, в силу леммы 5 равенство (10) не нарушится после применения подстановки s:

$$x_1 s = r(\rho_1 s, \ldots, \rho_n s, \sigma_1 s, \ldots, \sigma_n s).$$

Но $x_1 s = x_2$, $\sigma_i s = \sigma_i$ и по доказанному выше $\rho_1 s = \rho_1, \ldots, \rho_n s = \rho_n$. Следовательно,

$$x_2 = r(\rho_1, \ldots, \rho_n, \sigma_1, \ldots, \sigma_n),$$

т. е. $x_2 = x_1$, что противоречит независимости $x_1, \ldots, x_n$.

§ 17. Группа алгебраического уравнения

Доказанная в предыдущем параграфе теорема Руффини-Абеля обнаруживает только то, что для всех алгебраических уравнений данной степени $n \geqslant 5$ универсальной формулы решения в радикалах не существует. Но отсюда ещё не следует, что существуют уравнения с числовыми коэффициентами, которые нельзя решить в радикалах; ведь остаётся возможность того, что каждое уравнение может иметь своё особое решение в радикалах. Поэтому для более полного исследования вопроса о разрешимости алгебраических уравнений в радикалах нам придётся пойти дальше и изложить некоторые сведения, относящиеся к теории нормальных полей (полей Галуа).

Пусть

$$F(x) = A_0 x^n + \ldots + A_n = 0 \qquad (1)$$

— некоторое алгебраическое уравнение n-й степени над числовым полем P. Мы будем предполагать, что комплексные корни уравнения (1) различны [1]).

Присоединим к полю P корни $\alpha_1, \alpha_2, \ldots, \alpha_n$ уравнения (1). Мы получим поле $\Omega = P(\alpha_1, \alpha_2, \ldots, \alpha_n)$, носящее название *нормального поля* или поля Галуа относительно P. В частности, когда P есть область рациональности уравнения (1), то Ω называется просто нормальным полем или полем Галуа, слова «относительно P» опускаются (см. § 10, стр. 232).

Введём теперь весьма важное понятие группы уравнения. Обозначим через G совокупность всех таких подстановок симметрической группы S_n корней уравнения (1), которые, оставляя неподвижными элементы поля P, не нарушают ни одно рациональное соотношение между корнями $\alpha_1, \ldots, \alpha_n$ над полем P.

Покажем, что *множество G образует группу относительно умножения подстановок.*

Для этой цели воспользуемся следующим предложением, известным из теории групп: если M — *конечное* множество, то оно образует группу относительно алгебраической операции, определённой в этом множестве и подчиняющейся сочетательному закону.

Таким образом, нам надо показать, что умножение подстановок является для G алгебраической операцией, т. е. что умножение подстановок всегда выполнимо во множестве G.

Возьмём две произвольные подстановки s_1 и s_2 из G. Пусть подстановка s_1 переводит некоторое соотношение $r_1(\alpha_1, \ldots, \alpha_n) = 0$ между корнями уравнения (1) в соотношение $r_2(\alpha_1, \ldots, \alpha_n) = 0$, а подстановка s_2 переводит $r_2(\alpha_1, \ldots, \alpha_n) = 0$ в $r_3(\alpha_1, \ldots, \alpha_n) = 0$. Тогда произведение $s_1 s_2$ переведёт соотношение $r_1(\alpha_1, \ldots, \alpha_n) = 0$

---

[1]) В противном случае мы отделили бы кратные корни. Об отделении кратных корней см. в книге Куроша [1] или в § 29 книги Окунева [3].

в $r_3(\alpha_1, \ldots, \alpha_n) = 0$. Следовательно, произведение $s_1 s_2$ не нарушает ни одно из рациональных соотношений между корнями уравнения (1), в силу чего $s_1 s_2$ должно также принадлежать G. Так как G — конечное множество, то отсюда следует, что G образует группу.

Эта группа G носит название *группы Галуа* уравнения (1) над полем P или короче *группы уравнения* (1) над полем P. В тех случаях, когда P есть область рациональности уравнения, слова «над полем P» обычно опускаются.

Пример 1. Найдём группу квадратного уравнения

$$x^2 + px + q = 0 \tag{2}$$

с рациональными коэффициентами, имеющего два различных действительных иррациональных корня α_1 и α_2.

Очевидно, что в данном случае P есть поле рациональных чисел. Далее, всякое рациональное соотношение $r(\alpha_1, \alpha_2) = 0$ между корнями уравнения (2) можно предполагать целым рациональным, так как α_1 и α_2 являются алгебраическими числами (относительно поля рациональных чисел). Кроме того, мы можем предположить, что в соотношение $r(\alpha_1, \alpha_2) = 0$ каждый из корней α_1, α_2 входит со степенью, не превосходящей 1, так как в противном случае мы могли бы с помощью уравнения (2) понизить степень соответствующего корня. Таким образом, соотношение $r(\alpha_1, \alpha_2) = 0$ можно записать в виде

$$r(\alpha_1, \alpha_2) = a\alpha_1\alpha_2 + b\alpha_1 + c\alpha_2 + d = 0$$

(a, b, c, d — рациональные числа).

Но по формулам Вьета $\alpha_1\alpha_2 = q$. Следовательно, полагая $aq + d = m$, имеем:

$$r(\alpha_1, \alpha_2) = b\alpha_1 + c\alpha_2 + m = 0.$$

Так как $\alpha_2 = -p - \alpha_1$, то получается, что

$$r(\alpha_1, \alpha_2) = (b - c)\alpha_1 + (m - pc) = 0.$$

Если бы $b - c \neq 0$, то мы имели бы, что $\alpha_1 = \frac{pc - m}{b - c}$, а это невозможно, так как α_1 иррационально. Следовательно, $b - c = 0$, в силу чего соотношение $r(\alpha_1, \alpha_2) = 0$ принимает такой окончательный вид:

$$r(\alpha_1, \alpha_2) = b(\alpha_1 + \alpha_2) + m = 0. \tag{3}$$

Соотношение (3), очевидно, не нарушается (даже остаётся одним и тем же) при любой подстановке симметрической группы S_2 второй степени. Следовательно, S_2 и есть группа уравнения (2).

К понятию группы уравнения можно подойти и с несколько иной точки зрения. Назовём *автоморфизмом* нормального поля Ω

уравнения (1) относительно P такой изоморфизм Ω с самим собою, при котором каждому элементу a из P ставится в соответствие тот же самый элемент a, т. е. при котором элемент a остаётся без изменения. Условимся автоморфизмы в отличие от подстановок обозначать большими латинскими буквами, чаще всего последними буквами латинского алфавита. При этом если какой-нибудь автоморфизм S переводит элемент ω поля Ω в элемент ω', то мы будем записывать это обстоятельство в форме равенства

$$\omega S = \omega'.$$

Введём теперь понятие умножения автоморфизмов поля Ω. Пусть S_1 и S_2 — два произвольных автоморфизма поля Ω и пусть

$$\omega S_1 = \omega', \quad \omega' S_2 = \omega''.$$

Тогда соответствие $\omega \to \omega''$ будет взаимно однозначным и при этом соответствии $\omega_1 + \omega_2 \to \omega_1'' + \omega_2''$, $\omega_1\omega_2 \to \omega_1''\omega_2''$, а элементы из поля P не будут изменяться. Короче говоря, это соответствие будет автоморфизмом поля Ω; мы его обозначим через S_1S_2 и назовём *произведением автоморфизмов* S_1 и S_2.

Покажем, что *множество H автоморфизмов нормального поля образует относительно введённой операции умножения группу, совпадающую с точностью до изоморфизма с группой G уравнения* (1).

Возьмём какое-нибудь рациональное соотношение $r(\alpha_1, \ldots, \alpha_n) = 0$ над P между корнями уравнения (1) и посмотрим, что произойдёт с этим соотношением при автоморфизме S нормального поля Ω. Так как автоморфизм S сохраняет алгебраические операции поля Ω и не изменяет элементы поля P, то автоморфизм S переведёт соотношение $r(\alpha_1, \ldots, \alpha_n) = 0$ в соотношение $r(\alpha_1 S, \ldots, \alpha_n S) = 0$. Далее, так как α_i — корень уравнения (1), то $f(\alpha_i) = 0$. Очевидно, что автоморфизм S равенство $f(\alpha_i) = 0$ переведёт в $f(\alpha_i S) = 0$. Мы видим, что $\alpha_i S$ есть корень уравнения (1), т. е. $\alpha_i S = \alpha_{j_i}$. При этом, если $i \neq k$, то $\alpha_i S \neq \alpha_k S$, так как S является взаимно однозначным соответствием. Таким образом, автоморфизм S вызывает некоторую подстановку

$$s = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}$$

корней уравнения (1), переводящую соотношение $r(\alpha_1, \ldots, \alpha_n) = 0$ в соотношение $r(\alpha_{i_1}, \ldots, \alpha_{i_n}) = 0$. Иными словами, автоморфизм S вызывает подстановку s группы G уравнения (1). Поставим эту подстановку s в соответствие S:

$$S \to s. \tag{4}$$

Покажем, что соответствие (4) является изоморфизмом между множеством H и группой G.

Пусть T — ещё один автоморфизм поля Ω и пусть T соответствует та же подстановка s, что и автоморфизму S: $T \to s$. Любой элемент ω из Ω есть многочлен от $\alpha_1, \ldots, \alpha_n$ над P:

$$\omega = f(\alpha_1, \ldots, \alpha_n).$$

Поэтому

$$\omega S = f(\alpha_1 S, \ldots, \alpha_n S) = f(\alpha_{i_1}, \ldots, \alpha_{i_n}) = f(\alpha_1 T, \ldots, \alpha_n T) = \omega T.$$

Так как ω — произвольно, то отсюда вытекает, что $S = T$.

Легко, далее, убедиться, что для всякой подстановки s из G можно указать автоморфизм S нормального поля Ω такой, что $S \to s$. В самом деле, если

$$s = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

то подстановка s переведёт произвольный элемент $\omega = f(\alpha_1, \ldots, \alpha_n)$ из Ω в элемент $\omega' = f(\alpha_{i_1}, \ldots, \alpha_{i_n})$. При другом выражении $\omega = g(\alpha_1, \ldots, \alpha_n)$ элемента ω через корни уравнения (1) подстановка s переведёт ω в тот же самый элемент ω'. Это следует из того, что рациональное соотношение

$$f(\alpha_1, \ldots, \alpha_n) = g(\alpha_1, \ldots, \alpha_n)$$

между корнями уравнения (1) не должно нарушаться вследствие подстановки s группы G. Таким образом, подстановка s вызывает соответствие

$$\omega \to \omega', \tag{5}$$

не зависящее от способа выражения элемента ω через корни уравнения (1), причём, очевидно, что если a — элемент поля P, то $a \to a$. Возьмём теперь другой элемент θ из Ω. Если $\theta \to \omega'$, то обратная подстановка s^{-1} будет переводить один и тот же элемент ω' в различные элементы ω и θ, что невозможно, так как мы только что убедились, что всякая подстановка группы G, в частности s^{-1}, должна, независимо от способа выражения элемента ω' через корни уравнения (1), переводить ω' в один и тот же элемент. Стало быть, соответствие (5) является не только однозначным, но и взаимно однозначным. Наконец, для произвольного ω' можно указать такое ω, что $\omega \to \omega'$. Именно, таким элементом ω будет элемент, получающийся из ω' в результате применения обратной подстановки s^{-1}. Итак, мы убедились, что соответствие (5) есть не что иное, как взаимно однозначное отображение поля Ω на самого себя, оставляющее элементы поля P неподвижными.

Покажем, что соответствие (5) и будет искомым автоморфизмом S. Действительно, если $\omega \to \omega'$ и $\theta \to \theta'$, то подстановка s переведёт сумму $\omega + \theta = f(\alpha_1, \ldots, \alpha_n) + g(\alpha_1, \ldots, \alpha_n)$ в

$$f(\alpha_{i_1}, \ldots, \alpha_{i_n}) + g(\alpha_{i_1}, \ldots, \alpha_{i_n}) = \omega' + \theta',$$

в силу чего $\omega + \theta \to \omega' + \theta'$. Точно так же убеждаемся, что $\omega\theta \to \omega'\theta'$. Все эти рассуждения показывают, что соответствие (4) есть взаимно однозначное соответствие между множествами H и G.

Остаётся проверить, выполняются ли для соответствия (4) условия изоморфизма. Пусть

$$S \to s = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}; \quad T \to t = \begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix}.$$

Тогда для произвольного элемента $\omega = f(\alpha_1, \dots, \alpha_n)$ из Ω получаем:

$$\omega(ST) = (\omega S)\, T = f(\alpha_{i_1}, \dots, \alpha_{i_n})\, T = f(\alpha_{j_1}, \dots, \alpha_{j_n}) = \omega'.$$

Мы видим отсюда, что элемент ω' получился из элемента ω с помощью подстановки

$$\begin{pmatrix} 1 & 2 & \dots & n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} = st.$$

Следовательно, $ST \to st$.

Итак, H и G оказались изоморфными. Благодаря этому обстоятельству мы можем не отличать G от H и в случае надобности понимать под группой уравнения (1) группу автоморфизмов его нормального поля Ω. Так как группа автоморфизмов H зависит лишь от поля Ω, то мы будем иногда G (и H) называть группой нормального поля Ω.

Прежде чем переходить к дальнейшему изучению группы уравнения, отметим некоторые свойства нормального поля Ω и алгебраических расширений.

Теорема 27. *Всякий элемент ω нормального поля Ω является корнем некоторого многочлена, неприводимого над P.*

Доказательство. Поскольку ω — элемент поля Ω, он выражается в виде многочлена от корней уравнения (1) над P:

$$\omega = f(\alpha_1, \dots, \alpha_n).$$

Подвергая $f(\alpha_1, \dots, \alpha_n)$ всевозможным подстановкам симметрической группы S_n, мы получим $n!$ элементов поля Ω:

$$\theta_1 = \omega, \ \theta_2, \ \dots, \ \theta_{n!}.$$

Составим вспомогательное уравнение

$$g(x) = (x - \theta_1)(x - \theta_2) \dots (x - \theta_{n!}) = 0.$$

На основании основной теоремы о симметрических многочленах нетрудно установить, что $g(x)$ есть многочлен над P. Вместе с тем $g(x)$ имеет в качестве одного из корней ω. Очевидно, что ω также будет корнем одного из неприводимых (в P) множителей многочлена $g(x)$, и теорема доказана.

Теорема 28. *Если корень ω многочлена $p(x)$, неприводимого в P, лежит в нормальном поле Ω, то и все корни $p(x)$ лежат в Ω.*

Доказательство. Как и выше, строим вспомогательный многочлен $g(x)$ над P, имеющий ω своим корнем. С одной стороны, все корни многочлена $g(x)$ лежат в поле Ω. С другой стороны, $g(x)$ имеет с многочленом $p(x)$ общий корень ω и потому в силу неприводимости $p(x)$ делится на $p(x)$. Но в таком случае корни $p(x)$ должны войти в состав корней $g(x)$. Отсюда следует, что корни $p(x)$ должны лежать в Ω.

Теорема 29. *Если нормальное поле Ω изоморфно отображается на некоторое промежуточное поле Δ ($P \subseteq \Delta \subseteq \Omega$) так, что элементы P остаются неподвижными, то $\Delta = \Omega$.*

Доказательство. Прежде всего легко убедиться, что нормальное поле Ω образует векторное пространство над P относительно операций сложения элементов Ω и умножения элемента Ω на элемент поля P. В самом деле, любой элемент нормального поля $\Omega = P(\alpha_1, \ldots, \alpha_n)$ является многочленом от $\alpha_1, \ldots, \alpha_n$ с коэффициентами из P. Складывая два таких многочлена, мы, очевидно, получим снова многочлен от $\alpha_1, \ldots, \alpha_n$ с коэффициентами из P. Точно также умножая многочлен от $\alpha_1, \ldots, \alpha_n$ с коэффициентами из P на элемент поля P, мы получим снова многочлен от $\alpha_1, \ldots, \alpha_n$ с коэффициентами из P. Кроме того, выполняются и те требования, которые в определении векторного пространства были сформулированы в виде аксиом I — V (см. стр. 42—43). Очевидно, что промежуточное поле Δ относительно тех же операций будет подпространством пространства Ω.

Далее, легко показать, что Ω — конечномерное пространство. Действительно, поскольку α_i ($i = 1, 2, \ldots, n$) является корнем алгебраического уравнения (1) n-й степени, любая степень α_i^k с показателями $k \geqslant n$ будет линейно выражаться через низшие степени $\alpha_i^0 = 1, \alpha_i, \ldots, \alpha_i^{n-1}$. Поэтому каждый элемент a нормального поля Ω можно выразить в виде многочлена от $\alpha_1, \ldots, \alpha_n$ над P, содержащего $\alpha_1, \ldots, \alpha_n$ с показателями, не превосходящими $n - 1$. Таким образом совокупность произведений $\alpha_1^{k_j} \alpha_2^{l_j} \ldots \alpha_n^{s_j}$ с показателями $k_j, l_j, \ldots, s_j$, не превосходящими $n - 1$, образуют конечный базис пространства Ω, в силу чего Ω конечномерно.

Теперь мы можем доказать нашу теорему. В теории векторных пространств доказывается, что конечномерное пространство нельзя изоморфно отобразить на собственное подпространство так, чтобы элементы P оставались неподвижными. Следовательно, Δ не может быть собственным подполем Ω и потому $\Delta = \Omega$.

Введём весьма важное понятие продолжения изоморфизма. Оно понадобится при изучении дальнейших свойств группы уравнения.

Пусть K и $\bar{K}$ — два изоморфных кольца (числовых или нечисловых, безразлично), Δ — расширение кольца K, а $\bar{\Delta}$ — расширение кольца $\bar{K}$. Пусть Δ и $\bar{\Delta}$ также изоморфны. Мы назовём изоморфизм

$\Delta \cong \overline{\Delta}$ [1]) *продолжением изоморфизма* $K \cong \overline{K}$, если всякий элемент a кольца K, отображающийся при изоморфизме $K \cong \overline{K}$ на $\overline{a}$, при изоморфизме $\Delta \cong \overline{\Delta}$ попрежнему отображается на $\overline{a}$.

Пример 2. Возьмём в качестве K поле всех действительных чисел, а в качестве $\overline{K}$ — поле матриц вида

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix},$$

где a — произвольное действительное число.

Для поля K алгебраическими операциями будут, очевидно, обычные арифметические сложение и умножение, а для $\overline{K}$ — матричные сложение и умножение. Приведём в соответствие каждому действительному числу a матрицу $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$:

$$a \to \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

Нетрудно убедиться, что это соответствие является изоморфизмом: $K \cong \overline{K}$.

Затем в качестве расширения Δ возьмём поле комплексных чисел, а в качестве расширения $\overline{\Delta}$ — поле матриц вида

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \tag{6}$$

где a, b — произвольные действительные числа. Читатель может сам проверить, что множество матриц вида (6) в самом деле образует поле относительно операций сложения и умножения матриц.

Приведём в соответствие каждому комплексному числу $a + bi$ матрицу $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$:

$$a + bi \to \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \tag{7}$$

Нетрудно убедиться, что это соответствие также является изоморфизмом: $\Delta \cong \overline{\Delta}$.

Положим $b = 0$. Тогда соответствие (7) превратится в

$$a \to \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

---

[1]) $\cong$ есть сокращённое обозначение изоморфизма.

Мы видим, что при изоморфизме $\Delta \cong \bar{\Delta}$ действительное число a попрежнему отображается в матрицу

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

Следовательно, изоморфизм $\Delta \cong \bar{\Delta}$ есть продолжение изоморфизма $K \cong \bar{K}$.

Отметим следующие свойства продолжения изоморфизма:

Теорема 30. *Если P и $\bar{P}$ — изоморфные поля, то кольцо многочленов $P[x]$ можно изоморфно отобразить на кольцо многочленов $\bar{P}[x]$ так, чтобы изоморфизм $P[x] \cong \bar{P}[x]$ являлся продолжением изоморфизма $P \cong \bar{P}$.*

Доказательство. Пусть при изоморфизме $P \cong \bar{P}$ элемент a поля P отображается на элемент $\bar{a}$ поля $\bar{P}$: $a \to \bar{a}$. Тогда произвольному многочлену $f(x) = a_0 + a_1 x + \ldots + a_m x^m$ из $P[x]$ можно поставить в соответствие многочлен $\bar{f}(x) = \bar{a}_0 + \bar{a}_1 x + \ldots + \bar{a}_m x^m$ из $\bar{P}[x]$:

$$f(x) = a_0 + a_1 x + \ldots + a_m x^m \to \bar{f}(x) = \bar{a}_0 + \bar{a}_1 x + \ldots + \bar{a}_m x^m. \quad (8)$$

Читатель может без труда проверить, что (8) является взаимно однозначным соответствием между $P[x]$ и $\bar{P}[x]$. Покажем, что соответствие (8) является изоморфизмом колец $P[x]$ и $\bar{P}[x]$.

В самом деле, если $g(x) = b_0 + b_1 x + \ldots + b_l x^l$ — ещё один произвольный многочлен из $P[x]$, то

$$g(x) = b_0 + b_1 x + \ldots + b_l x^l \to \bar{g}(x) = \bar{b}_0 + \bar{b}_1 x + \ldots + \bar{b}_l x^l.$$

Пусть, например, $l \leqslant m$. Тогда

$$f(x) + g(x) = c_0 + c_1 x + \ldots + c_m x^m,$$

где $c_i = a_i + b_i$, причём в случае $l < m$ надо положить $b_{l+1} = \ldots = b_m = 0$. Отсюда

$$f(x) + g(x) \to \bar{c}_0 + \bar{c}_1 x + \ldots + \bar{c}_m x^m = \bar{f}(x) + \bar{g}(x).$$

Аналогично можно убедиться, что $f(x)\,g(x) \to \bar{f}(x)\,\bar{g}(x)$.

Остаётся показать, что изоморфизм $P[x] \cong \bar{P}[x]$ есть продолжение изоморфизма $P \cong \bar{P}$.

Обращаемся к многочлену $f(x) = a$, где a — элемент P. Для него соответствие (8) принимает вид $a \to \bar{a}$. Мы видим, что при изоморфизме $P[x] \cong \bar{P}[x]$ элемент a поля P попрежнему отображается на элемент $\bar{a}$ поля $\bar{P}$.

Переходя ко второму свойству продолжения изоморфизма, отметим, что если $f(x)$ — некоторый многочлен над полем P, то под

$\bar{f}(x)$ мы будем всё время подразумевать такой многочлен над изоморфным полем $\bar{P}$, на который отображается $f(x)$ при изоморфизме $P[x] \cong \bar{P}[x]$, упомянутом в теореме 30.

Отметим ещё, что если $p(x)$ — многочлен, неприводимый над P, то согласно теореме 30 многочлен $\bar{p}(x)$ также неприводим над $\bar{P}$.

Теорема 31. *Если P и $\bar{P}$ — изоморфные числовые поля, θ — корень многочлена $p(x)$, неприводимого над P, и $\bar{\theta}$ — корень многочлена $\bar{p}(x)$, то изоморфизм $P \cong \bar{P}$ можно продолжить до изоморфизма $P(\theta) \cong \bar{P}(\bar{\theta})$, при котором θ будет отображаться на $\bar{\theta}$.*

Доказательство. Пусть степень многочлена $p(x)$ равна k. Тогда произвольный элемент γ алгебраического расширения $P(\theta)$ будет единственным образом выражаться в виде

$$\gamma = a_0 + a_1\theta + \ldots + a_{k-1}\theta^{k-1} \qquad (a_i \text{ — элементы из } P)$$

(см. замечание на стр. 229). Пусть при изоморфизме $P \cong \bar{P}$ элемент a_i отображается на $\bar{a}_i$. Поставим элементу γ в соответствие элемент $\bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \ldots + \bar{a}_{k-1}\bar{\theta}^{k-1}$ поля $\bar{P}$:

$$\gamma = a_0 + a_1\theta + \ldots + a_{k-1}\theta^{k-1} \to \bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \ldots + \bar{a}_{k-1}\bar{\theta}^{k-1}. \quad (9)$$

Покажем, что соответствие (9) является изоморфизмом $P(\theta) \cong \bar{P}(\bar{\theta})$.

Пусть какой-нибудь элемент δ расширения $P(\theta)$ отображается на тот же самый элемент, что и γ: $\delta \to \bar{\gamma}$. Тогда, если $\delta = b_0 + b_1\theta + \ldots + b_{k-1}\theta^{k-1}$, то

$$\delta \to \bar{b}_0 + \bar{b}_1\bar{\theta} + \ldots + \bar{b}_{k-1}\bar{\theta}^{k-1} = \bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \ldots + \bar{a}_{k-1}\bar{\theta}^{k-1}.$$

Но многочлен $\bar{p}(x)$ неприводим над P. Следовательно, $\bar{\gamma}$ должно единственным образом выражаться в виде многочлена от $\bar{\theta}$ степени, не превосходящей $k-1$, и с коэффициентами из $\bar{P}$. В силу этого $\bar{a}_0 = \bar{b}_0$, $\bar{a}_1 = \bar{b}_1$, $\ldots$, $\bar{a}_{k-1} = \bar{b}_{k-1}$. Отсюда благодаря изоморфизму $P \cong \bar{P}$ получается, что $a_0 = b_0$, $a_1 = b_1$, $\ldots$, $a_{k-1} = b_{k-1}$, т. е. $\delta = \gamma$.

Очевидно, что для всякого элемента $\bar{\gamma}$ из $\bar{P}(\bar{\theta})$ можно указать в $P(\theta)$ такой элемент γ, которому $\bar{\gamma}$ и ставится в соответствие.

Всё это вместе взятое означает, что (9) есть взаимно однозначное соответствие между $P(\theta)$ и $\bar{P}(\bar{\theta})$.

Возьмём теперь из расширения $P(\theta)$ два произвольных элемента

$$\gamma_1 = a_0 + a_1\theta + \ldots + a_{k-1}\theta^{k-1} \quad \text{и} \quad \gamma_2 = b_0 + b_1\theta + \ldots + b_{k-1}\theta^{k-1}.$$

Этим элементам в алгебраическом расширении $\bar{P}(\bar{\theta})$ будут соответствовать

$$\bar{\gamma}_1 = \bar{a}_0 + \bar{a}_1\bar{\theta} + \ldots + \bar{a}_{k-1}\bar{\theta}^{k-1}, \quad \bar{\gamma}_2 = \bar{b}_0 + \bar{b}_1\bar{\theta} + \ldots + \bar{b}_{k-1}\bar{\theta}^{k-1}.$$

Отсюда сумме

$$\gamma_1+\gamma_2=(a_0+b_0)+(a_1+b_1)\theta+\ldots+(a_{k-1}+b_{k-1})\theta^{k-1}$$

будет соответствовать

$$(\bar{a}_0+\bar{b}_0)+(\bar{a}_1+\bar{b}_1)\bar{\theta}+\ldots+(\bar{a}_{k-1}+\bar{b}_{k-1})\bar{\theta}^{k-1}=\bar{\gamma}_1+\bar{\gamma}_2.$$

Точно так же убеждаемся, что произведению $\gamma_1\gamma_2$ должно соответствовать произведение $\bar{\gamma}_1\bar{\gamma}_2$. Мы видим, что соответствие (9) есть изоморфизм $P(\theta)\cong\bar{P}(\bar{\theta})$.

Покажем, что изоморфизм $P(\theta)\cong\bar{P}(\bar{\theta})$ является продолжением изоморфизма $P\cong\bar{P}$. Действительно, если γ — элемент поля P, то $\gamma=a_0$, и для такого элемента a_0 соответствие (9) принимает вид $a_0\to\bar{a}_0$. Таким образом, при изоморфизме $P(\theta)\cong\bar{P}(\bar{\theta})$ элементы поля P отображаются совершенно так же, как и при изоморфизме $P\cong\bar{P}$.

Наконец, соответствие (9) будет, очевидно, переводить элемент θ в $\bar{\theta}$.

Вернёмся к группе уравнения (1). Мы собираемся рассмотреть следующие свойства группы уравнения.

Мы знаем, что всякий элемент ω нормального поля Ω является корнем некоторого многочлена $p(x)$, неприводимого над P (см. теорему 27). Назовём два элемента ω и ω' поля Ω *сопряжёнными*, если они являются корнями одного и того же многочлена $p(x)$ неприводимого над P. Тогда можно высказать такую теорему:

Теорема 32. *Всякая подстановка (автоморфизм) группы уравнения* (1) *переводит элемент* ω *нормального поля* Ω *в сопряжённый элемент* ω'. *Обратно, если* ω' *— элемент, сопряжённый* ω, *то в группе уравнения* (1) *существует по меньшей мере одна подстановка (автоморфизм), переводящая* ω *в* ω'.

Доказательство. Пусть ω является корнем многочлена $p(x)$, неприводимого над P, и s — произвольная подстановка из группы G уравнения (1). Равенство $p(\omega)=0$ есть рациональное соотношение над P между корнями $\alpha_1,\ldots,\alpha_n$ уравнения (1), так как ω рационально выражается через $\alpha_1,\ldots,\alpha_n$. Следовательно, применяя к этому соотношению подстановку s, получаем на основании определения группы уравнения, что $p(\omega s)=0$. Таким образом, $\omega'=\omega s$ есть корень того же самого многочлена $p(x)$, т. е. элемент ω' сопряжён с ω.

Обратно, пусть ω' — некоторый элемент, сопряжённый ω. Это значит, что ω и ω' являются корнями одного и того же многочлена $p(x)$, неприводимого над P. Воспользуемся теперь теоремой 31: возьмём в качестве $\bar{P}$ то же самое поле P, а в качестве изоморфизма $P\cong\bar{P}$ будем рассматривать соответствие $a\to a$, оставляющее элементы a поля P неподвижными [1]). Тогда на основании теоремы 31

[1]) Так называемый тождественный изоморфизм поля P с самим собой.

этот изоморфизм $P \cong \bar{P}$ можно продолжить до изоморфизма $P(\omega) \cong P(\omega')$, переводящего ω в ω'. Если $P(\omega) = \Omega$, то в силу теоремы 29 расширение $P(\omega')$ должно также совпадать с Ω, и мы имеем автоморфизм поля Ω, переводящий ω в ω'.

Если же $P(\omega)$ не совпадает с Ω, а является лишь частью Ω, то рассуждаем следующим образом. Возьмём из Ω элемент θ, не лежащий в $P(\omega)$. Обозначим через $p_1(x)$ неприводимый над P многочлен, имеющий θ своим корнем. Над полем $P(\omega)$ многочлен $p_1(x)$ может оказаться приводимым. Пусть $p_1(x)$ над полем $P(\omega)$ следующим образом разлагается в произведение неприводимых множителей:

$$p_1(x) = q_1(x)\, q_2(x) \ldots q_r(x) \tag{10}$$

(множители $q_i(x)$ будут выше первой степени). Предположим для определённости, что θ — корень $q_1(x)$. На основании теоремы 30 разложению (10) будет над полем $P(\omega')$ соответствовать разложение

$$p_1(x) = \bar{q}_1(x)\, \bar{q}_2(x) \ldots \bar{q}_r(x)$$

на множители $\bar{q}_i(x)$, неприводимые в $P(\omega')$. По теореме 28 все корни многочлена $p_1(x)$ должны лежать в нормальном поле Ω. Тем самым в Ω будут лежать и все корни множителя $\bar{q}_1(x)$. Пусть θ' — один из корней $q_1(x)$. Тогда по теореме 31

$$P(\omega, \theta) \cong P(\omega', \theta'), \tag{11}$$

причём изоморфизм (11) есть продолжение изоморфизма $P(\omega) \cong P(\omega')$. Если $P(\omega, \theta) = \Omega$, то по теореме 29 будет и $P(\omega', \theta') = \Omega$; вследствие этого (11) будет искомым автоморфизмом нормального поля Ω, переводящим ω в ω'. Если же $P(\omega, \theta)$ не совпадает с Ω, а является лишь частью Ω, то продолжаем наш процесс. В конечном счёте мы, очевидно, придём к искомому автоморфизму поля Ω [1]).

Условимся в дальнейшем говорить, что элемент ω нормального поля Ω допускает подстановки (автоморфизмы) группы поля Ω, если ω остаётся неизменным при любой подстановке группы. Тогда из теоремы 32 вытекает такое важное.

Следствие. *Элемент ω нормального поля Ω тогда и только тогда допускает подстановки группы G поля Ω, когда ω лежит в поле P.*

---

[1]) Конечность этого процесса вытекает из следующих соображений. Каждое расширение $P(\omega)$, $P(\omega, \theta)$, ... можно рассматривать как подпространство конечномерного пространства Ω. Но в теории векторных пространств доказывается, что последовательность подпространств конечномерного пространства, в которой каждое подпространство содержится в следующем, не может быть бесконечной (в связи с ограниченностью размерности подпространств).

В самом деле, пусть ω лежит в поле P. Тогда, очевидно, ω не меняется при любой подстановке из группы G.

Обратно, если ω не меняется при любой подстановке из группы G, то по только что доказанной теореме 32 все элементы, сопряжённые с ω, должны совпадать с ω. Но последнее возможно лишь тогда, когда ω есть корень многочлена $p(x)$ первой степени: $p(x)=x-a$, где a — элемент P.

Таким образом, $\omega-a=0$ или $\omega=a$. Мы видим, что ω оказался элементом поля P.

Существенную роль в нашем изложении будет играть следующая

Теорема 33. *Пусть Ω — нормальное поле и G — группа уравнения* (1) *над P. Тогда всякому промежуточному полю P' ($P\subseteq P'\subseteq\Omega$) соответствует подгруппа G' группы G, которая в свою очередь является группой уравнения* (1), *но уже над P', а именно G' является совокупностью таких подстановок из G, которые оставляют неподвижным любой элемент P'. При этом поле P' определяется подгруппой G' однозначно: P' есть совокупность всех элементов Ω, «допускающих» подстановки из G', т. е. остающихся неподвижными при этих подстановках.*

Доказательство. Группа G' уравнения (1), рассматриваемого над P', есть, очевидно, совокупность таких подстановок корней $\alpha_1, \ldots, \alpha_n$ уравнения, которые не нарушают ни одно рациональное соотношение между $\alpha_1, \ldots, \alpha_n$ над P' и оставляют элементы P' неподвижными. Отсюда подстановка s из G' и подавно не будет нарушать рациональное соотношение между $\alpha_1, \ldots, \alpha_n$ над P, так как эти соотношения можно рассматривать и как соотношения между $\alpha_1, \ldots, \alpha_n$ над P'. Следовательно, s есть элемент группы G, откуда G' есть подгруппа G (возможно совпадающая с G).

Покажем теперь, что G' состоит из всех тех подстановок группы G, которые оставляют элементы P' неподвижными. Обозначим совокупность таких подстановок через G''. Ясно, что группа G' будет содержаться в G'': $G'\subseteq G''$. Кроме того, легко видеть, что G'' является также группой, так как произведение двух подстановок, оставляющих элементы поля P' неподвижными, также оставляет неподвижными элементы P'.

Пусть теперь t — некоторая подстановка из G''. Эта подстановка не нарушает ни одного рационального соотношения между $\alpha_1, \ldots, \alpha_n$ над P, так как t содержится в группе G уравнения (1). Рассмотрим некоторое рациональное соотношение $r(\alpha_1, \ldots, \alpha_n)=0$ над P'. С одной стороны, это соотношение не нарушается подстановкой t, так как, выражая коэффициенты соотношения через корни $\alpha_1, \ldots, \alpha_n$ уравнения (1), мы получим соотношение между $\alpha_1, \ldots, \alpha_n$ над P. С другой стороны, коэффициенты соотношения $r(\alpha_1, \ldots, \alpha_n)=0$ подстановкой t не меняются. Следовательно, t содержится в G', откуда $G''\subseteq G'$. Сопоставляя $G''\subseteq G'$ и $G'\subseteq G''$, видим, что $G''=G'$.

Для завершения доказательства теоремы остаётся убедиться в однозначности определения поля P' посредством подгруппы G'.

Пусть ω — некоторый элемент Ω, допускающий подстановки из G'. Тогда согласно следствию из теоремы 32 сразу вытекает, что ω есть элемент P'.

Для дальнейшего введём понятие гомоморфизма групп. Изоморфизм является частным случаем этого понятия.

Пусть некоторая группа H_1 однозначно отображается на группу H_2, причём отображение может и не быть взаимно однозначным. Мы назовём отображение *гомоморфным,* если произведению любых двух элементов группы H_1 соответствует произведение соответствующих элементов группы H_2. Мы будем гомоморфное отображение группы H_1 на группу H_2 записывать в виде: $H_1 \backsim H_2$.

Теперь мы можем высказать следующую теорему:

Теорема 34. *Если поле P', промежуточное между P и нормальным полем Ω ($P \subseteq P' \subseteq \Omega$), есть в свою очередь нормальное поле некоторого многочлена $g(x)$ над P, то группа G поля Ω над P гомоморфно отображается на группу $\bar{G}$ поля P' над P.*

Доказательство. Обозначим корни $g(x)$ через $\beta_1, \beta_2, \ldots, \beta_m$. Каждая подстановка s из $\bar{G}$ будет перемещать корень β_i в корень β_j. При этом различные корни β_i и β_k будут перемещаться подстановкой s в различные корни. В самом деле, если бы $\beta_i s$ было равно $\beta_k s$ при $\beta_i \neq \beta_k$, то, применяя к равенству $\beta_i s = \beta_k s$ обратную подстановку s^{-1}, мы нарушили бы соотношение $\beta_i s = \beta_k s$, а именно получили бы $\beta_i \neq \beta_k$. Таким образом, получилось бы противоречие с тем, что s^{-1} принадлежит группе G.

Итак, подстановка s вызывает некоторую подстановку

$$\bar{s} = \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \beta_m \\ \beta_{i_1} & \beta_{i_2} & \ldots & \beta_{i_m} \end{pmatrix}$$

корней многочлена $g(x)$. При этом, очевидно, $\bar{s}$ не будет нарушать ни одно рациональное соотношение $r(\beta_1, \ldots, \beta_m) = 0$ над P и будет оставлять неподвижными элементы из P, т. е. $\bar{s}$ будет подстановкой группы поля P' над P.

Обратно, всякую подстановку $\bar{s}$ из группы поля P' над P можно с помощью процесса, использованного при доказательстве теоремы 32, продолжить до подстановки (автоморфизма) s группы поля Ω над P.

Поставим теперь подстановке s в соответствие подстановку $\bar{s}$, вызываемую s:

$$s \to \bar{s}. \tag{12}$$

Легко видеть, что соответствие (12) будет отображать произведение $s_1 s_2$ в произведение $\bar{s}_1 \bar{s}_2$, вследствие чего соответствие (12) есть гомоморфное отображение группы G на группу $\bar{G}$.

§ 18. Уравнения с симметрической группой

В этом параграфе мы покажем, что уравнения степени $n \geqq 5$ с симметрической группой неразрешимы в радикалах. Предварительно докажем несколько теорем.

Теорема 35. *Если первообразный корень нечётной n-й степени из единицы не принадлежит полю P, то группа двучленного уравнения*

$$f(x) = x^n - 1 = 0 \tag{1}$$

над P является коммутативной.

Доказательство. Пусть ε — первообразный корень уравнения (1). Тогда $P(\varepsilon)$ есть нормальное поле уравнения (1), так как все корни уравнения (1) рационально выражаются через ε, а именно являются степенями ε. Обозначим через $p(x)$ тот из неприводимых над P множителей многочлена $f(x)$, который имеет ε своим корнем. Очевидно, что $P(\varepsilon)$ будет нормальным полем и для $p(x)$, вследствие чего группа $p(x)$ совпадает с группой уравнения (1). Пусть $\theta_1 = \varepsilon$, $\theta_2 = \varepsilon^{k_2}, \ldots, \theta_m = \varepsilon^{k_m}$ — все корни многочлена $p(x)$. По теореме 31 $P(\theta_1) \cong P(\theta_i)$, причём этот изоморфизм является продолжением тождественного изоморфизма $P \cong P$ (т. е. изоморфизма, оставляющего неподвижными элементы поля P) и переводит θ_1 в θ_i. Так как, очевидно, $P(\theta_i) \subseteq P(\theta_1)$, то в силу теоремы 29 получается, что $P(\theta_1) = P(\theta_i)$. Мы видим отсюда, что $P(\theta_1) \cong P(\theta_i)$ есть автоморфизм нормального поля $P(\theta_1)$ над P, т. е. один из элементов группы G уравнения (1) над P. Таким образом, группа G должна состоять из m различных подстановок $s_1 = 1, s_2, \ldots, s_m$:

$$\theta_1 s_1 = \theta_1, \quad \theta_1 s_2 = \theta_2, \ldots, \theta_1 s_m = \theta_m.$$

Найдём, чему равны $\theta_1(s_i s_j)$ и $\theta_1(s_j s_i)$. Имеем:

$$\theta_1(s_i s_j) = \theta_i s_j = (\varepsilon^{k_i}) s_j = (\varepsilon s_j)^{k_i} = (\theta_1 s_j)^{k_i} = \theta_j^{k_i} = \varepsilon^{k_i k_j},$$

$$\theta_1(s_j s_i) = \varepsilon^{k_i k_j}.$$

Отсюда $s_i s_j = s_j s_i$, т. е. G есть коммутативная группа.

Теорема 36. *Если*

$$F(x) = A_0 x^n + A_1 x^{n-1} + \ldots + A_n = 0 \tag{2}$$

есть уравнение с симметрической группой над полем P, то многочлен $F(x)$ неприводим над P.

Доказательство. Пусть $\alpha_1, \ldots, \alpha_n$ — корни уравнения (2) и пусть многочлен $F(x)$ приводим над P:

$$F(x) = c p_1^{k_1}(x) p_2^{k_2}(x) \ldots p_r^{k_r}(x),$$

где $p_i(x)$ — различные неприводимые над P многочлены. Пусть α_i есть корень $p_1(x)$, а α_j — корень $p_2(x)$. Возьмём подстановку, перемещающую только корни α_i и α_j:

$$s=\begin{pmatrix}1 & 2 & \dots & i & j & \dots & n\\ 1 & 2 & \dots & j & i & \dots & n\end{pmatrix},$$

и применим её к рациональному соотношению $p_1(\alpha_i)=0$. Получим $p_1(\alpha_j)=0$. Отсюда многочлен $p_1(x)$ должен в силу неприводимости делиться на $p_2(x)$, вследствие чего $p_1(x)$ и $p_2(x)$ совпадают с точностью до множителя нулевой степени, что невозможно.

Теорема 37. *Если* (2) *есть попрежнему уравнение с симметрической группой над полем P и степень уравнения* (2) *больше двух, то ни один из корней такого уравнения не может быть рациональной функцией первообразного корня k-й (нечётной) степени из единицы над P.*

Иными словами, ни один из корней α_i уравнения (2) не лежит в поле $P(\varepsilon)$, где ε — первообразный корень k-й степени из единицы.

Доказательство. Согласно теореме 35 группа G' поля $P(\varepsilon)$ над P является коммутативной. Так как по теореме 36 многочлен $F(x)$ неприводим над P, то, предполагая, что в $P(\varepsilon)$ лежит по меньшей мере один корень $F(x)$, получаем на основании теоремы 28, что в $P(\varepsilon)$ должны лежать все корни $F(x)$, т. е. $P(\alpha_1, \dots, \alpha_n)\subseteq P(\varepsilon)$. Таким образом, на основании теоремы 34 получаем, что $G'\curvearrowright G$, где G — группа уравнения (2) над P или, что то же, поля Ω над P. Из гомоморфизма $G'\curvearrowright G$ вытекает, что G — также коммутативная группа [1]). Но последнее невозможно, так как по условию G — симметрическая группа: $G=S_n$, а при $n>2$ симметрическая группа не является коммутативной.

Теорема 38. *Пусть попрежнему* (2) *есть уравнение степени $n>2$ с симметрической группой над P. Тогда группа уравнения* (2) *над полем $K=P(\varepsilon_1, \dots, \varepsilon_k)$, где ε_i — первообразные корни из единицы простой нечётной степени p_i (p_i между собой различны), будет также симметрической.*

Доказательство. Обозначим для краткости $P(\alpha_1, \dots, \alpha_n)$ через Ω и $K(\alpha_1, \dots, \alpha_n)$ через Σ. Мы утверждаем, что $K=P(\varepsilon)$, где ε — первообразный корень $p_1p_2\dots p_k$-й степени из единицы. В самом деле, с одной стороны,

$$P(\varepsilon)\subseteq P(\varepsilon_1, \dots, \varepsilon_k),$$

так как нетрудно убедиться, что произведение $\varepsilon_1\varepsilon_2\dots\varepsilon_k$ есть первообразный корень $p_1p_2\dots p_k$-й степени из единицы. С другой стороны,

$$P(\varepsilon_1, \dots, \varepsilon_k)\subseteq P(\varepsilon),$$

[1]) Действительно, если элементы $a'b'$ группы G', отображаются на элементы a, b группы G, то $a'b'\to ab$ и $b'a'\to ba$. Так как группа G' коммутативная, то $a'b'=b'a'$. Отсюда и $ab=ba$, т. е. группа G также должна быть коммутативной.

так как каждое из ε_i является корнем двучленного уравнения $x^m - 1 = 0$, где $m = p_1 p_2 \dots p_k$, и потому является некоторой степенью ε. Отсюда следует, что $P(\varepsilon_1, \dots, \varepsilon_k) = P(\varepsilon)$.

На основании теоремы 37 имеем, что $\Sigma \supset K$, но не равно K. Относительно полей P и K представляются две возможности: либо $K = P$, либо $K \supset P$. Если $K = P$, то теорема очевидна. Поэтому пусть

$$P \subset K \subset \Sigma. \tag{3}$$

По теореме 33 соотношению (3) должно соответствовать соотношение

$$G_1 \supseteq G_2 \supseteq E,$$

где G_1 — группа поля Σ над P, G_2 — группа того же поля над K и E — единичная группа (т. е. подгруппа, состоящая только из единицы). Очевидно, что G_2 есть вместе с тем и группа уравнения (2) над K.

Поле Σ можно, очевидно, рассматривать как нормальное поле многочлена

$$h(x) = F(x)(x^{m-1} + x^{m-2} + \dots + x + 1) \qquad (m = p_1 p_2 \dots p_k)$$

над P. Таким образом, G_1 есть, кроме того, группа уравнения $h(x) = 0$ над P.

Обозначим корни многочлена $x^{m-1} + \dots + x + 1$ через θ_1, θ_2, $\dots$, θ_{m-1}. Тогда корнями $h(x)$ будут: $\alpha_1, \alpha_2, \dots, \alpha_n, \theta_1, \dots, \theta_{m-1}$. Эти корни между собой различны, так как по теореме 37 многочлены $F(x)$ и $x^{m-1} + \dots + x + 1$ не могут иметь общих корней.

Посмотрим, что представляет собою подстановка s группы G_1. Эта подстановка не может перемещать α_i в θ_j. В самом деле, если бы $\alpha_i s$ было равно θ_j, то $F(\alpha_i) = 0$ под влиянием s перешло бы в $F(\theta_j) = 0$, что в силу теоремы 37 невозможно. Таким образом, подстановка s из G_1 должна иметь вид

$$s = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n & \theta_1 & \theta_2 & \dots & \theta_{m-1} \\ \alpha_{i_1} & \alpha_{i_2} & \dots & \alpha_{i_n} & \theta_{j_1} & \theta_{j_2} & \dots & \theta_{j_{m-1}} \end{pmatrix}. \tag{4}$$

Покажем, что индексы $i_1, i_2, \dots, i_n$ во второй строчке выражения (4) образуют произвольную перестановку n чисел $1, 2, \dots, n$. Действительно, так как $\Sigma \supseteq \Omega \supseteq P$ и группа поля Ω над P есть симметрическая группа S_n, то по теореме 34 $G_1 \backsim S_n$ и при этом гомоморфизме

$$s \to \bar{s} = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_{i_1} & \alpha_{i_2} & \dots & \alpha_{i_n} \end{pmatrix}. \tag{5}$$

Теперь, если бы в выражении (4) индексы $i_1, i_2, \dots, i_n$ не пробегали бы всех $n!$ перестановок из n чисел $1, 2, \dots, n$, то соответствие (5) не исчерпывало бы всех подстановок $\bar{s}$ симметрической

группы S_n, вследствие чего G_1 не могла бы гомоморфно отображаться на S_n (она отображалась бы на правильную часть S_n).

Посмотрим, далее, что представляет собою группа G_2 поля Σ над K. Согласно теореме 33 группа G_2 состоит из тех и только тех подстановок G_1, которые оставляют неподвижными элементы промежуточного поля K и, в частности, оставляют неподвижными θ_j. Отсюда получаем, исходя из выражения (4), что группа G_2 должна состоять из всевозможных подстановок t вида

$$t=\begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n & \theta_1 & \theta_2 & \dots & \theta_{m-1}\\ \alpha_{i_1} & \alpha_{i_2} & \dots & \alpha_{i_n} & \theta_1 & \theta_2 & \dots & \theta_{m-1}\end{pmatrix}.$$

Теперь без труда убеждаемся, что группы G_2 и S_n изоморфны. Именно, изоморфным отображением G_2 на S_n здесь будет соответствие

$$t\to\bar{s}=\begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n\\ \alpha_{i_1} & \alpha_{i_2} & \dots & \alpha_{i_n}\end{pmatrix}.$$

Итак, теорема доказана: группа G_2 уравнения (2) над K с точностью до изоморфизма совпала с симметрической группой S_n.

Мы подошли вплотную к основной теореме о неразрешимости в радикалах алгебраических уравнений с симметрической группой.

Теорема 39. *Если группа уравнения n-й степени*

$$F(x)=A_0x^n+A_1x^{n-1}+\dots+A_n=0 \qquad (6)$$

(A_i — комплексные числа) *является симметрической и степень уравнения больше четырёх, то уравнение неразрешимо в радикалах.*

Доказательство. Роль поля P здесь играет область рациональности Δ уравнения $K=\Delta(\varepsilon_1,\dots,\varepsilon_k)$. По теореме 38 группа уравнения (6) будет симметрической и над полем K. Таким образом, к этой группе будут также принадлежать подстановки

$$s=\begin{pmatrix}1&2&3&4&5&6&\dots&n\\ 2&3&4&5&1&6&\dots&n\end{pmatrix},\quad t=\begin{pmatrix}1&2&3&4&5&6&\dots&n\\ 1&2&4&5&3&6&\dots&n\end{pmatrix},$$

$$u=\begin{pmatrix}1&2&3&4&5&\dots&n\\ 2&3&1&4&5&\dots&n\end{pmatrix},$$

Эти подстановки, как и любые подстановки группы уравнения над K, не будут изменять элементы K и, в частности, будут оставлять неподвижными ε_i. Поэтому мы можем дословно повторить все рассуждения, приведённые в доказательстве теоремы Руффини-Абеля (см. § 16). В результате мы получим, что корни α_1 и α_2 уравнения (6) равны, что невозможно в силу неприводимости многочлена $F(x)$.

Для завершения нашего изложения остаётся привести конкретный пример уравнения степени $n \geqslant 5$ с симметрической группой.

Для этой цели отметим некоторые факты из теории групп подстановок, не вдаваясь в детали.

Пусть G — некоторая группа подстановок n-й степени. Возможны два случая: 1) в группе G существуют подстановки, переводящие число 1 в любое заданное число k ($k = 1, 2, \ldots, n$); 2) подстановки группы G перемещают 1 не во всякое заданное число k.

В первом случае группа G называется *транзитивной*, а во втором случае — *интранзитивной*.

Роль транзитивной группы в теории алгебраических уравнений видна из следующей теоремы:

Теорема 40. *Если уравнение n-й степени*

$F(x) = A_0x^n + A_1x^{n-1} + \ldots + A_n = 0$ (A_i — комплексные числа) *неприводимо над своей областью рациональности Δ, то группа такого уравнения транзитивна.*

Доказательство. Если многочлен $F(x)$ неприводим над полем Δ, то по теореме 32 в группе уравнения наверное найдутся подстановки, переводящие корень α_1 в любой другой корень α_k уравнения $F(x) = 0$, чем транзитивность группы и обнаруживается.

Мы воспользуемся следующим свойством транзитивной группы: *транзитивная группа простой степени p, содержащая транспозицию, является симметрической группой.*

Примечание. *Транспозицией* называется подстановка, перемещающая только два числа, т. е. подстановка вида

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & i & j & \ldots & n \\ 1 & 2 & 3 & \ldots & j & i & \ldots & n \end{pmatrix}.$$

Обычно транспозиция, перемещающая числа i, j, обозначается сокращённо символом (i, j).

Пользуясь этим свойством транзитивной группы, мы сейчас докажем такую теорему:

Теорема 41. *Всякое уравнение простой степени $p \geqslant 5$ с рациональными коэффициентами, неприводимое над полем рациональных чисел и имеющее лишь одну пару чисто комплексных корней, неразрешимо в радикалах.*

Доказательство. По теореме 40 группа такого уравнения $F(x) = 0$ есть транзитивная группа. Она к тому же является группой простой (p-й) степени.

Пусть $\alpha_1 = a + bi$ и $\alpha_2 = a - bi$ — чисто комплексные корни уравнения; остальные корни $\alpha_3, \alpha_4, \ldots, \alpha_p$ согласно условиям теоремы должны быть действительными.

Рассмотрим какое-нибудь рациональное соотношение

$$f(\alpha_1, \alpha_2, \ldots, \alpha_p) = 0 \qquad (7)$$

между корнями $\alpha_1, \alpha_2, \ldots, \alpha_p$ над полем рациональных чисел. Это соотношение можно даже считать относительно $\alpha_1, \ldots, \alpha_n$ целым рациональным. Подставляя в левую часть равенства (7) вместо α_1 и α_2 соответственно $a+bi$ и $a-bi$ и собирая отдельно действительные и мнимые члены, получаем:

$r(\alpha_1, \alpha_2, \ldots, \alpha_p)=P+iQ=0$ (P, Q — действительные числа), откуда $P=Q=0$. Применим теперь к $r(\alpha_1, \alpha_2, \ldots, \alpha_p)$ транспозицию (12). В силу сопряжённости корней α_1 и α_2 это равносильно перемене знаков мнимых частей выражения $r(\alpha_1, \alpha_2, \ldots, \alpha_p)$:

$$r(\alpha_2, \alpha_1, \ldots, \alpha_p)=P-iQ.$$

Но по доказанному выше $P=Q=0$. Следовательно,

$$r(\alpha_2, \alpha_1, \ldots, \alpha_p)=0.$$

Таким образом, транспозиция (12) не нарушила соотношения (7). Это означает, что транспозиция (12) содержится в группе G уравнения. Отсюда согласно вышеупомянутому свойству транзитивной группы вытекает, что группа G является симметрической: $G=S_n$. Значит, по теореме 39 уравнение $F(x)=0$ неразрешимо в радикалах.

Из этой теоремы сразу получается, что уравнения пятой степени вида

$$x^5-q^2x-q=0, \tag{8}$$

где q — простое число, неразрешимы в радикалах.

В самом деле, неприводимость уравнения (8) сразу обнаруживается с помощью критерия Эйзенштейна. Пользуясь теоремой Штурма[1]), можно установить, что уравнение (8) имеет всего три действительных корня. Таким образом, уравнение (8) удовлетворяет всем условиям теоремы 41 и потому не может быть решено в радикалах.

Изучение свойств алгебраических расширений с помощью аппарата теории групп является основной задачей теории Галуа. Рассмотренный здесь вопрос о разрешимости алгебраических уравнений в радикалах есть одно из приложений этой теории. В работах советских математиков (С. О. Шатуновский, Н. Г. Чеботарёв, Б. Н. Делоне и др.) теория Галуа получила дальнейшее обобщение и развитие[2]). Н. Г. Чеботарёв посвятил ей две свои монографии: «Основы теории Галуа», ч. 1 (ГТТИ, 1934) и ч. 2 (ОНТИ, 1937)

---

[1]) См. стр. 324, А. П. Доморяд, Численные и графические методы решения уравнений.

[2]) Более подробные сведения можно найти в сборнике «Математика в СССР за 30 лет» (Гостехиздат, 1948).

и «Теория Галуа» (ОНТИ, 1936). Им же получены фундаментальные результаты, относящиеся к одному из разделов теории Галуа, так называемой «Проблеме резольвент». За свои исследования в этой области (основные результаты этих исследований опубликованы в статье «Проблема резольвент», Юбилейный сборник АН СССР, 1947) Н. Г. Чеботарёв удостоен Сталинской премии первой степени. Б. Н. Делоне разработал оригинальную геометрическую теорию, представляющую собой обобщение теории Галуа. Б. Н. Делоне и Д. К. Фаддеев применили геометрические методы к решению наиболее трудных задач теории Галуа.

§ 19. О разрешимости алгебраических уравнений в квадратных радикалах

Существует целый ряд геометрических задач на построение, сводящихся к нахождению корней некоторого алгебраического уравнения n-й степени

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0. \qquad (I)$$

К таким задачам относятся, например, задача удвоения куба, задача трисекции угла.

В курсе теории геометрических построений доказывается, что некоторое выражение α тогда и только тогда может быть построено с помощью циркуля и линейки, когда оно получается в результате решения уравнений не выше второй степени. Например, выражение

$$\alpha = \sqrt[4]{1 + \sqrt{2}}$$

можно построить с помощью циркуля и линейки, так как оно получается в результате решения ряда уравнений не выше второй степени. А именно, $\alpha_1 = \sqrt{2}$ есть корень квадратного уравнения $x^2 - 2 = 0$, и мы можем построить $\sqrt{2}$ с помощью циркуля и линейки. Далее, $\alpha_2 = 1 + \sqrt{2}$ является корнем уравнения первой степени $x - (1 + \alpha_1) = 0$, и мы можем $1 + \sqrt{2}$ также построить с помощью циркуля и линейки — придётся складывать отрезки, соответствующие 1 и $\sqrt{2}$. Затем $\alpha_3 = \sqrt{\alpha_2}$ является корнем квадратного уравнения $x^2 - \alpha_2 = 0$, и, поскольку α_2 было уже заранее построено, мы этот корень без труда построим с помощью циркуля и линейки. Наконец, $\alpha = \sqrt{\alpha_3}$ есть корень квадратного уравнения, $x^2 - \alpha_3 = 0$, а так как α_3 уже построено, то мы построим с помощью циркуля и линейки и $\alpha = \sqrt{\alpha_3}$.

Таким образом, возникает следующий естественный вопрос — при каких условиях, необходимых и достаточных, уравнение (I) решается в квадратных радикалах?

Введём предварительно важное понятие конечного расширения. Пусть Δ и Ω — два каких-нибудь числовых поля, причём пусть Ω является расширением Δ. Мы будем часто пользоваться обозначением $\Delta \subset \Omega$, выражающим, что Δ есть подполе Ω (Ω есть расширение Δ). Назовём систему k элементов $\alpha_1, \ldots, \alpha_k$ из Ω линейно зависимой относительно Δ, если в Δ можно выбрать такие числа $c_1, \ldots, c_k$, не все равные нулю, чтобы

$$c_1\alpha_1 + \ldots + c_k\alpha_k = 0. \tag{1}$$

В противном случае, когда равенство (1) имеет место только при $c_1 = c_2 = \ldots = c_n = 0$, система элементов $\alpha_1, \ldots, \alpha_k$ называется линейно независимой.

Понятие конечного расширения мы теперь определим следующим образом.

Назовём расширение Ω поля Δ *конечным* относительно Δ, если в Ω существует такая линейно независимая система элементов $\omega_1, \ldots, \omega_s$, что всякий элемент α из Ω является линейной комбинацией $\omega_1, \ldots, \omega_s$:

$$\alpha = a_1\omega_1 + \ldots + a_s\omega_s,$$

где a_i — элементы из Δ.

Систему элементов $\omega_1, \ldots, \omega_s$ мы будем называть *базисом* конечного расширения Ω. Вообще мы назовём базисом всякую линейно независимую систему элементов из Ω, через которую линейно выражается любой элемент расширения.

Отметим следующие основные свойства конечного расширения.

1°. *Количество элементов базиса конечного расширения Ω не зависит от выбора базиса.*

Мы будем это количество элементов базиса называть степенью конечного расширения Ω относительно Δ и будем обозначать через $(\Omega : \Delta)$.

2°. *Если $(\Omega : \Delta) = s$, то всякая система $s+1$ элементов конечного расширения Ω линейно зависима* (относительно Δ).

3°. *Если $(\Omega : \Delta) = s$, то каждый элемент α конечного расширения Ω является алгебраическим относительно Δ, а именно является корнем многочлена над Δ не выше s-й степени.*

Свойство 3° легко вытекает из свойства 2°. В самом деле, в силу свойства 2° система $s+1$ элементов $\alpha^0 = 1, \alpha, \alpha^2, \ldots, \alpha^s$ должна быть линейно зависима (относительно Δ), т. е.

$$c_0 + c_1\alpha + c_2\alpha^2 + \ldots + c_s\alpha^s = 0, \tag{2}$$

где c_i — элементы Δ, причём среди c_i по меньшей мере один эле-

мент отличен от нуля. Но равенство (2) как раз свидетельствует о том, что α есть корень многочлена

$$f(x)=c_0+c_1x+\ldots+c_sx^s$$

над Δ не выше s-й степени.

Приведём примеры конечных расширений.

Пример 1. Легко видеть, что поле K комплексных чисел является конечным расширением поля D действительных чисел и притом второй степени относительно D.

В самом деле, 1 и $i=\sqrt{-1}$ образуют линейно независимую систему относительно D, так как равенство $c\cdot 1+d\cdot i=0$, где c и d — действительные числа, имеет место лишь при $c=d=0$. Кроме того, всякое комплексное число, как известно, выражается в виде $a+bi$, где a, b — действительные числа. Следовательно, 1, i является базисом K.

Так как базис 1, i состоит из двух элементов, то степень $(K:D)$ равна двум.

Пример 2. Алгебраическое расширение $P(\theta)$, получающееся путём присоединения к полю P корня θ многочлена $p(x)$ k-й степени, неприводимого в P, является конечным расширением k-й степени поля P.

Действительно, 1, θ, $\theta^2,\ldots,$ θ^{k-1} образуют линейно независимую систему: если бы имело место равенство

$$c_0+c_1\theta+c_2\theta^2+\ldots+c_{k-1}\theta^{k-1}=0,$$

где c_i — числа из P, не равные одновременно нулю, то θ было бы корнем уравнения ниже k-й степени с коэффициентами из P, что в силу неприводимости $p(x)$ невозможно.

Затем, как известно, всякий элемент алгебраического расширения $P(\theta)$ следующим образом выражается через θ:

$$\alpha=a_0+a_1\theta+a_2\theta^2+\ldots+a_{k-1}\theta^{k-1}\quad(a_i\text{ — числа из }P).$$

Мы видим отсюда, что $P(\theta)$ есть конечное расширение P с базисом 1, θ, $\theta^2,\ldots,$ θ^{k-1}. Очевидно, что степень $(P(\theta):P)$ равна k, так как базис 1, θ, $\theta^2,\ldots,$ θ^{k-1} состоит из k элементов.

В теории конечных расширений основную роль играют следующие две теоремы:

Теорема 42. *Если* Ω_1 *— конечное расширение поля* Δ, *а* Ω_2 *— конечное расширение* Ω_1, *то* Ω_2 *есть конечное расширение* Δ *и* $(\Omega_2:\Delta)=(\Omega_2:\Omega_1)(\Omega_1:\Delta)$.

Теорема 43[1]). *Пусть* Ω *— конечное расширение поля* Δ. *Тогда всякое расширение* Σ *поля* Δ, *содержащееся в* Ω, *будет*

[1]) Доказательство теорем 42 и 43, а также выводы других свойств конечных расширений можно найти в книге: Л. Я. Окунев, Высшая алгебра, глава одиннадцатая (см. § 53—56 в издании 1944 г. или § 52—55 в издании 1949 г.).

конечным расширением Δ. *Степень* $(\Sigma:\Delta)$ *будет при этом делителем степени* $(\Omega:\Delta)$.

Из теоремы 42 вытекает, что всякое расширение $P(\theta, \ldots, \theta_n)$, получающееся в результате присоединения к P элементов $\theta_1, \ldots, \theta_n$, алгебраических относительно P, является конечным расширением поля P. Отсюда, в частности, следует, что нормальное поле $\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$ многочлена $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ n-й степени является конечным расширением области рациональности Δ многочлена $f(x)$.

Теперь мы в состоянии дать ответ на вопрос, поставленный в начале параграфа. А именно, можно высказать следующую теорему:

Теорема о разрешимости алгебраического уравнения в квадратных радикалах[1]). *Пусть* $f(x)$ *— многочлен* n*-й степени,* Δ *— область рациональности и* Ω *— нормальное поле многочлена* $f(x)$. *Для того чтобы уравнение* $f(x) = 0$ *было разрешимо в квадратных радикалах, необходимо и достаточно, чтобы* $(\Omega:\Delta) = 2^m$.

Неразрешимость уравнения $f(x) = 0$ в квадратных радикалах не исключает, однако, того обстоятельства, что некоторые корни уравнения всё же могут выражаться через квадратные радикалы. Но если многочлен $f(x)$ неприводим в своей области рациональности Δ, то справедлива следующая

Теорема 44. *Если многочлен* $f(x)$ *степени* $n \geqslant 2$ *неприводим в своей области рациональности* Δ *и какой-нибудь его корень выражается через квадратные радикалы, то и все остальные корни многочлена выражаются через квадратные радикалы.*

Доказательство. Пусть корень α многочлена $f(x)$ выражается через квадратные радикалы $\rho_1 = \sqrt{A_1}$, $\rho_2 = \sqrt{A_2}$, $\rho_3 = \sqrt{A_3}, \ldots, \rho_k = \sqrt{A_k}$, где A_1 — элемент Δ, A_2 — элемент $\Delta(\rho_1)$, A_3 — элемент $\Delta(\rho_1, \rho_2), \ldots, A_k$ — элемент $\Delta(\rho_1, \rho_2, \ldots, \rho_{k-1})$.

Первый радикал можно рассматривать как корень многочлена

$$\varphi_1(x) = x^2 - A_1$$

над Δ. Этот многочлен $\varphi_1(x)$, кроме ρ_1, имеет корнем другой квадратный радикал $\rho_1' = -\rho_1$.

Так как A_2 — элемент поля $\Delta(\rho_1)$, то $A_2 = a + b\rho_1$, где a, b — элементы Δ. Отсюда следует, что второй радикал ρ_2 можно рассматривать как корень многочлена

$$\varphi_2(x) = (x^2 - a - b\rho_1)(x^2 - a + b\rho_1) = x^4 - 2ax^2 + (a^2 + b^2A_1)$$

над Δ. Этот многочлен, кроме ρ_2, имеет своими корнями квадратные радикалы $\rho_2' = -\rho_2$, $\rho_2'' = \sqrt{a - b\rho_1}$, $\rho_2''' = -\rho_2''$.

---

[1]) Доказательство этой теоремы см. в кн.: Л. Окунев, Высшая алгебра, § 55, издания 1944 г. или § 54 издания 1949 г.

Затем, так как A_3 — элемент поля $\Delta(\rho_1, \rho_2)$, то $A_3 = c + d\rho_2$, причём c и d — элементы $\Delta(\rho_1)$, т. е. $c = m + n\rho_1$ и $d = m_1 + n_1\rho_1$, где m, n, m_1 и n_1 — элементы Δ. Отсюда следует, что радикал ρ_3 можно рассматривать как корень многочлена над Δ

$$\varphi_3(x) = (x^2 - c - d\rho_2)(x^2 - c + d\rho_2)(x^2 - \bar{c} - \bar{d}\bar{\rho}_2)(x^2 - \bar{c} + \bar{d}\bar{\rho}_2),$$

где

$$\bar{c} = m - n\rho_1, \quad \bar{d} = m_1 - n_1\rho_1, \quad \bar{\rho}_2 = \sqrt{a - b\rho_1}.$$

Многочлен $\varphi_3(x)$, кроме ρ_3, будет иметь в качестве корней и другие квадратные радикалы.

Аналогичным образом можно убедиться, что ρ_4 будет корнем многочлена $\varphi_4(x)$ над Δ, имеющего, кроме ρ_4, в качестве корней и другие квадратные радикалы, и т. д.

Составим теперь многочлен

$$F(x) = \varphi_1(x)\varphi_2(x)\ldots\varphi_k(x)$$

над Δ. Присоединяя корни многочлена $F(x)$ к Δ, получим для $F(x)$ нормальное поле Ω. В этом поле, очевидно, будет лежать и корень α данного многочлена $f(x)$. Так как многочлен $f(x)$ неприводим в Δ, то по известному свойству нормального поля (см. теорему 28) все корни $f(x)$ должны лежать в Ω. Тем самым все корни многочлена $f(x)$ выражаются через корни $F(x)$, т. е. через квадратные радикалы, что и требовалось доказать.

§ 20. О разрешимости в квадратных радикалах уравнений 3-й и 4-й степеней

В теории геометрических построений чаще всего приходится иметь дело с задачами, сводящимися к алгебраическим уравнениям не выше четвёртой степени. Так как корни квадратного уравнения, очевидно, выражаются через квадратные радикалы, то остаётся рассмотреть случаи уравнений третьей и четвёртой степеней.

Относительно уравнения третьей степени имеет место

Теорема 45. *Уравнение третьей степени*

$$f(x) = a_0x^3 + a_1x^2 + a_2x + a_3 = 0 \qquad (1)$$

тогда и только тогда разрешимо в квадратных радикалах, когда многочлен $f(x) = a_0x^3 + a_1x^2 + a_2x + a_3$ *приводим в своей области рациональности* Δ.

Доказательство. Пусть многочлен $f(x)$ неприводим в Δ. Тогда, присоединяя к Δ какой-нибудь корень θ уравнения (1), мы получим конечное расширение $\Sigma = \Delta(\theta)$ третьей степени относительно Δ. Это Σ, очевидно, будет содержаться в нормальном поле Ω многочлена $f(x)$. По теореме 43 степень $(\Omega : \Delta)$ должна делиться на степень $(\Sigma : \Delta) = 3$. Поэтому степень $(\Omega : \Delta)$ не может равняться

2^m, откуда следует, что уравнение (1) не может быть решено в квадратных радикалах.

Обратно, если многочлен $f(x)$ приводим в Δ, то

$$f(x)=(x-a)(x^2+px+q),$$

где a, p и q — числа из Δ. Мы видим отсюда, что уравнение (1) распадается на уравнения первой и второй степени и тем самым решается в квадратных радикалах.

На основании теоремы 44 мы можем утверждать, что если многочлен $f(x)=a_0x^3+a_1x^2+a_2x+a_3$ третьей степени неприводим в своей области рациональности Δ, то ни один из его корней не выражается в квадратных радикалах.

Отметим одно следствие, вытекающее из теоремы 45.

Следствие. *Уравнение третьей степени с рациональными коэффициентами тогда и только тогда разрешимо в квадратных радикалах, когда оно имеет по меньшей мере один рациональный корень.*

В самом деле, в этом случае область рациональности Δ многочлена $f(x)=a_0x^3+a_1x^2+a_2x+a_3$ будет совпадать с полем рациональных чисел. Затем, как известно, многочлен третьей степени с рациональными коэффициентами тогда и только тогда неприводим в поле рациональных чисел, когда этот многочлен не имеет рациональных корней. Отсюда следствие становится очевидным.

Впрочем, в справедливости следствия можно убедиться и независимо от теоремы 45. Приводим здесь соответствующие рассуждения.

Доказательство. Без ограничения общности выводов можно предположить, что старший коэффициент уравнения третьей степени с рациональными коэффициентами равен единице:

$$f(x)=x^3+ax^2+bx+c=0. \qquad (1')$$

Если уравнение (1′) имеет рациональный корень, то оно, очевидно, разрешимо в квадратных радикалах.

Обратно, пусть уравнение (1′) разрешимо в квадратных радикалах. Допустим, что при этом уравнение (1′) не имеет рациональных корней. Возьмём какой-нибудь из корней x_1 уравнения или, что то же, многочлена $f(x)$. Пусть x_1 выражается через квадратные радикалы: $\rho_1=\sqrt{A_1}$, $\rho_2=\sqrt{A_2}$, ..., $\rho_k=\sqrt{A_k}$, где $k\geqslant 1$, A_1 — элемент поля Δ рациональных чисел, A_2 — элемент поля $\Delta(\rho_1)$ и т. д. Мы вправе предположить, что радикал ρ_k не лежит в $\Delta(\rho_1,\ldots,\rho_{k-1})$; в противном случае он был бы лишним; в частности, ни один из радикалов не должен лежать в поле рациональных чисел. Кроме того, мы имеем право допустить, что x_1 не лежит в

поле $\Delta(\rho_1, \ldots, \rho_{k-1})$; если бы это было не так, то радикал ρ_k снова был бы лишним. Таким образом, мы можем написать, что

$$x_1 = p + q\rho_k,$$

где p, q — элементы поля $\Delta(\rho_1, \ldots, \rho_{k-1})$ и $q \neq 0$.

Несложный подсчёт показывает, что $x_2 = p - q\rho_k$ также является корнем многочлена $f(x)$. А именно, подставляя значение x_1 в уравнение (1'), после очевидных преобразований получаем:

$$f(x_1) = P + Q\rho_k = 0,$$

где

$$P = p^3 + 3pq^2A_k + ap^2 + aq^2A_k + bp + c,$$

$$Q = 3p^2q + A_kq^3 + 2apq + bq.$$

Очевидно, что P и Q являются элементами поля $\Delta(\rho_1, \ldots, \rho_{k-1})$.

Теперь, если допустить, что $Q \neq 0$, то $\rho_k = -\frac{P}{Q}$ и радикал ρ_k лежал бы в $\Delta(\rho_1 \ldots, \rho_{k-1})$, что невозможно. Следовательно, $Q = 0$ и тем самым $P = 0$.

Подставляя в многочлен $f(x)$ вместо x число $x_2 = p - q\rho_k$, получаем после аналогичных преобразований, что

$$f(x_2) = P - Q\rho_k.$$

Но $P = Q = 0$. Значит, $f(x_2) = 0$, т. е. x_2 также является корнем многочлена $f(x)$. Отметим, что $x_2 \neq x_1$, так как $q \neq 0$.

Обозначим, далее, через x_3 третий корень многочлена $f(x)$. Тогда по формулам Вьета

$$x_1 + x_2 + x_3 = -a,$$

откуда

$$x_3 = -a - x_1 - x_2,$$

или, подставляя значения x_1 и x_2:

$$x_3 = -a - 2p.$$

Мы видим отсюда, что x_3 лежит в поле $\Delta(\rho_1, \ldots, \rho_{k-1})$, так как a и p лежат в $\Delta(\rho_1, \ldots, \rho_{k-1})$ [1]).

---

[1]) Применяя подобные рассуждения к уравнению 3-й степени с произвольными (не обязательно рациональными) коэффициентами, мы получим также другое доказательство теоремы 45.

Повторим для корня x_3 те же рассуждения, что и для x_1. Так как x_3 выражается через радикалы $\rho_1, \ldots, \rho_{k-1}$, то мы получим, что x_1 и x_2 лежат в $\Delta(\rho_1, \ldots, \rho_{k-2})$. Но это невозможно, так как x_1 не лежит уже в $\Delta(\rho_1, \ldots, \rho_{k-1})$.

Применим изложенное к следующим задачам на геометрические построения.

Задача об удвоении куба, как известно, сводится к уравнению третьей степени

$$f(x) = x^3 - 2 = 0 \tag{2}$$

с рациональными коэффициентами. Это уравнение, однако, не имеет рациональных корней. Следовательно, уравнение (2) нельзя решить в квадратных радикалах, и потому задача об удвоении куба неразрешима с помощью циркуля и линейки.

Обратимся теперь к знаменитой задаче о трисекции угла. Она заключается в следующем. Пусть дан угол α; требуется его разделить на три равные части.

Посмотрим, к какому алгебраическому уравнению можно свести эту задачу.

Обозначим искомый угол через φ. Тогда

$$\cos\alpha = \cos 3\varphi = 4\cos^3\varphi - 3\cos\varphi.$$

Поскольку угол α дан, мы можем считать его косинус также заданным. Поэтому полагаем $\cos\alpha = \frac{b}{2}$, а $\cos\varphi$ обозначаем через $\frac{x}{2}$. Таким образом,

$$\frac{b}{2} = 4\left(\frac{x}{2}\right)^3 - 3\left(\frac{x}{2}\right),$$

или окончательно:

$$f(x) = x^3 - 3x - b = 0.$$

Возьмём, например, $\alpha = \frac{\pi}{3}$. В этом случае $b = 1$, и мы получаем уравнение

$$x^3 - 3x - 1 = 0$$

с рациональными коэффициентами. Легко убедиться, что это уравнение не имеет рациональных корней. Стало быть, угол $\alpha = \frac{\pi}{3}$ нельзя разделить на три равные части с помощью циркуля и линейки.

Относительно уравнений четвёртой степени справедлива следующая

Теорема 46. *Пусть*

$$f(x) = x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 = 0 \tag{3}$$

— уравнение четвёртой степени. Если кубическая резольвента

этого уравнения имеет в Δ *хотя бы один корень, то корни уравнения* (3) *выражаются через квадратные радикалы. Обратно, если корни уравнения* (3) *выражаются через квадратные радикалы, то резольвента имеет по меньшей мере один корень в* Δ.

Доказательство. Пусть кубическая резольвента уравнения (3) имеет в Δ корень y_0. Тогда на основании известных преобразований (см. § 5) получаем

$$f(x)=\left(x^2+\frac{1}{2}ax+y_0\right)^2-(Ax+B)^2,$$

где

$$A^2=\frac{a^2}{4}-b+2y_0,\quad B^2=y_0^2-d,\quad 2AB=ay_0-c.$$

Таким образом, корни уравнения (3) будут выражаться через квадратные радикалы, возникающие при решении квадратных уравнений

$$x^2+\left(\frac{a}{2}+A\right)x+(y_0+B)=0, \tag{4}$$

$$x^2+\left(\frac{a}{2}-A\right)x+(y_0-B)=0. \tag{5}$$

Обратно, пусть корни уравнения (3) выражаются через квадратные радикалы. Тогда, обозначая через x_1, x_2 корни уравнения (4) и через x_3, x_4 — корни уравнения (5), получаем:

$$x_1x_2=y_0+B,\quad x_3x_4=y_0-B.$$

Отсюда находим, что

$$y_0=\frac{1}{2}(x_1x_2+x_3x_4).$$

Но корни уравнения (3) все выражаются через квадратные радикалы. Следовательно, y_0 также выражается через квадратные радикалы. Таким образом, если

$$y^3+ay^2+by+c=0 \tag{6}$$

— кубическая резольвента уравнения (3), то в силу произвольности y_0 все её корни выражаются через квадратные радикалы.

Итак, кубическая резольвента (6) разрешима в квадратных радикалах, а потому она по теореме 45 должна иметь по меньшей мере один корень в Δ.

Следствие. *Уравнение четвёртой степени с рациональными коэффициентами тогда и только тогда разрешимо в квадратных радикалах, когда его кубическая резольвента имеет по меньшей мере один рациональный корень.*

Теорему 46 можно дополнить следующим предложением.

Теорема 47. *Если многочлен четвёртой степени*

$$f(x)=x^4+a_1x^3+a_2x^2+a_3x+a_4$$

не имеет корней в своей области рациональности Δ *и один из корней* α *многочлена* $f(x)$ *выражается через квадратные радикалы, то и все остальные корни многочлена выражаются через квадратные радикалы.*

Доказательство. Если многочлен $f(x)$ приводим в своей области рациональности Δ, то он должен разлагаться в произведение двух квадратных многочленов с коэффициентами из Δ:

$$f(x)=(x^2+px+q)(x^2+p_1x+q_1),$$

так как по условию $f(x)$ не имеет в Δ корней. Таким образом, корни $f(x)$ найдутся из квадратных уравнений

$$x^2+px+q=0, \quad x^2+p_1x+q_1=0,$$

т. е. будут выражаться через квадратные радикалы.

Если многочлен $f(x)$ неприводим в своей области рациональности, то по теореме 44 все его корни должны выражаться через квадратные радикалы.

Пример. Через точку $A(2, 1)$ плоскости P провести на этой плоскости прямую g так, чтобы оси Ox и Oy прямоугольной системы координат Oxy на плоскости P отсекали на прямой g отрезок c длиною, равной пяти единицам выбранного масштаба.

Уравнение искомой прямой можно написать в следующем виде:

$$\frac{x}{u+2}+\frac{y}{v}=1.$$

Так как прямая g проходит через точку A (2, 1), то

$$\frac{2}{u+2}+\frac{1}{v}=1.$$

Отсюда

$$v=\frac{u+2}{u}.$$

Таким образом, подставляя это выражение v в уравнение

$$(u+2)^2+v^2=25,$$

получаем после некоторых преобразований:

$$u^4+4u^3-20u^2+4u+4=0. \tag{7}$$

Уравнение (7) не имеет рациональных корней; следовательно, по теореме 47 возможно лишь одно из двух: либо все корни уравнения (7) выражаются через квадратные радикалы, либо ни один из корней уравнения не выражается через квадратные радикалы.

Резольвентою уравнения (7) является

$$y^3+10y^2-50=0.$$

Нетрудно убедиться, что резольвента не имеет рациональных корней. Принимая во внимание теорему 46, мы заключаем отсюда, что ни один из корней уравнения (7) не выражается через квадратные радикалы. Следовательно, искомая прямая g не может быть построена с помощью циркуля и линейки.

Литература

1. Кузьмин Р. О. и Фаддеев Д. К., Алгебра и арифметика комплексных чисел, Учпедгиз, 1939.

В первой части книги содержатся сведения из теории комплексных чисел и излагается ряд приложений комплексных чисел, обнаруживающих плодотворпость введения их в науку. Вторая часть книги посвящена арифметике целых комплексных чисел.

2. Курош А. Г., Курс высшей алгебры, Гостехиздат, 1946 и 1950.

Учебник для университетов и педагогических институтов.

3. Маркушевич А. И., Деление с остатком в арифметике и алгебре, Издательство Академии педагогических наук, 1949.

Книга посвящена вопросу деления с остатком в произвольном кольце, удовлетворяющем некоторым условиям и, в частности, делению с остатком в кольце целых чисел и в кольце многочленов.

4. Окунев Л. Я., Высшая алгебра, Гостехиздат, 1944 и 1949.

Учебник для университетов и педагогических институтов.

5. Окунев Л. Я., Основы современной алгебры, Учпедгиз, 1941.

Книга посвящена элементам теории групп, теории колец и полей, а также теории разрешимости алгебраических уравнений в радикалах.

6. Проскуряков И. В., Числа и многочлены, Издательство Академии педагогических наук, 1949.

Целью книги является строгое определение чисел, многочленов и алгебраических дробей и обоснование их свойств, уже известных из средней школы.

7. Шапиро Г. М., Высшая алгебра, Учпедгиз, 1938.

Учебник для педагогических институтов.

А. П. ДОМОРЯД

ЧИСЛЕННЫЕ И ГРАФИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ

*

ВВЕДЕНИЕ

В средней школе изучаются главным образом такие уравнения и системы уравнений, которые допускают «точное» решение: для линейных уравнений и их систем корни рационально выражаются через коэффициенты, а в других случаях дело сводится, как правило, к квадратным или к двучленным уравнениям, и корни выражаются через коэффициенты с помощью радикалов.

Однако нередко в школу проникают «со стороны» и более сложные уравнения вроде

$$2^x - 4x = 0; \tag{1}$$

$$\left.\begin{aligned} x^2 + y &= 31, \\ y^2 + x &= 41 \end{aligned}\right\} \tag{2}$$

и т. п., причём ставится требование «элементарного» решения подобного рода уравнений.

Нельзя отрицать, что поиски искусственных приёмов решения таких уравнений могут принести некоторую пользу, но, во-первых, они дают чаще всего лишь некоторые из возможных решений и, кроме того, чрезмерное увлечение искусственными приёмами и особенно противопоставление «элементарных» решений решениям «неэлементарным» безусловно может принести большой вред, неверно ориентируя учащихся в тех требованиях, которые практика предъявляет математике.

Значительно полезнее в таких случаях познакомить учащихся с существованием общих приёмов решения уравнений, приёмов, позволяющих с любой степенью точности вычислить любой действительный корень как алгебраического, так и трансцендентного уравнения. Следует отметить, что эти общие приёмы нельзя считать менее законными, чем, например, решение уравнений в радикалах.

Но не только «со стороны» могут проникнуть в школу уравнения высших степеней и уравнения трансцендентные. Целый ряд задач элементарной математики приводит естественно к таким уравнениям.

Укажем для примера несколько таких задач.

Задача 1. Найти центральный угол сектора, площадь которого делится пополам хордой, стягивающей концы дуги [1]).

Задача 2. Найти дугу, длина которой равнялась бы длине линии косинуса [2]).

Считая центральный угол равным x радианам, легко придти в этих задачах, соответственно, к уравнениям

$$x - 2\sin x = 0, \tag{3}$$

$$x - \cos x = 0. \tag{4}$$

Задача 3. (Задача Архимеда.) На каком расстоянии от центра шара проходит секущая плоскость, отсекающая от шара $\frac{1}{n}$ часть его объёма.

Легко убедиться, что искомое расстояние удовлетворяет уравнению третьей степени:

$$x^3 - 3R^2x + 2\left(1 - \frac{2}{n}\right)R^3 = 0. \tag{5}$$

Задача 4. (О трисекции угла.) Разделить данный угол φ на три равные части.

Если, решая эту задачу алгебраическим способом, обозначить при $R = 1$

$$\cos\varphi = a, \quad \cos\frac{\varphi}{3} = x,$$

то для x будем иметь уравнение

$$4x^3 - 3x - a = 0, \tag{6}$$

вытекающее из известной формулы для косинуса тройного угла.

Задача 5. (Построение правильных многоугольников.) Построение правильных семиугольника, девятиугольника и одиннадцатиугольника связано с построением отрезков $2\cos\frac{2\pi}{7}$, $2\cos\frac{2\pi}{9}$, $2\cos\frac{2\pi}{11}$, удовлетворяющих соответственно (при $R = 1$) уравнениям

$$x^3 + x^2 - 2x - 1 = 0, \tag{7}$$

$$x^3 - 3x + 1 = 0, \tag{8}$$

$$x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1 = 0. \tag{9}$$

Нам кажется, что рассмотрению подобного рода задач следует время от времени уделять в школе некоторое внимание, так как составление уравнений является хорошим упражнением на программный материал; разъяснение же весьма несложного табличного способа решения уравнений, пригодного для любого уравнения вида

$$f(x) = 0 \tag{10}$$

с непрерывной функцией в левой части, даст учащимся ясное представление о практическом решении любых уравнений и будет способствовать приближению преподавания математики к нуждам прак-

---

[1]) Л. Эйлер, Introductio in Analysin infinitorum, 1748, т. II, п° 532.

[2]) Там же, п° 531.

тики. Кроме того, решение уравнений представляет источник получения иррациональных чисел «не радикального типа» и естественно приводит к объяснению сущности понятия трансцендентного числа.

Способ вычисления действительных корней уравнения (10), который в дальнейшем мы будем называть *табличным способом,* заключается в следующем: придавая аргументу x целочисленные значения, находим, например, что

$$f(a)<0; \quad f(a+1)>0.$$

Следовательно, при каком-то значении аргумента, лежащем между a и $a+1$, $f(x)$ обращается в нуль.

Придавая затем аргументу x значения $a+\frac{1}{10}$, $a+\frac{2}{10}$, $a+\frac{3}{10}$ и т. д., можно уточнить цифру десятых долей искомого корня, затем перейти к уточнению цифры сотых и т. д.

Конечно, табличный способ очень трудоёмок, и существуют способы, дающие гораздо быстрее значения действительных корней уравнений с большой точностью. Некоторые методы позволяют вычислять и мнимые корни уравнений. Следует подчеркнуть, что мнимые корни алгебраических уравнений играют большую роль во многих вопросах; так, например, в механике для «устойчивости движения» необходимо, чтобы действительные части мнимых корней некоторого уравнения были отрицательными.

Ещё чаще приходят к уравнениям, требующим применения численных методов, при решении различных задач механики, высшей математики, астрономии и т. д. Ограничимся одним лишь примером: Леверье при его колоссальных вычислениях, приведших к открытию Нептуна, пришлось столкнуться с решением уравнения [1])

$$3447z^6+14\,560z^5+22\,430z^4+25\,857z^3+$$
$$+29\,193z^2+11\,596z+5602=0.$$

На практике иногда приходится иметь дело с системами уравнений, в которых число уравнений превышает число неизвестных (например, в геодезии); довольно часто приходится решать уравнения, коэффициенты которых — приближённые числа, следовательно, точное решение уравнения не даёт точного решения задачи.

Влияние подобного рода причин на погрешность корней изучается в «приближённых вычислениях».

В дальнейшем мы будем предполагать, что коэффициенты решаемых уравнений — точные числа, а в системах уравнений число неизвестных совпадает с числом уравнений.

---

[1]) П. В. Мелентьев. Несколько новых методов и приёмов приближённых вычислений, 1937, стр. 26.

Настоящая статья посвящена рассмотрению методов, пригодных для решения любых уравнений с числовыми коэффициентами независимо от возможности решить их по специальным формулам (например, уравнения третьей степени) или с помощью каких-нибудь упрощающих приёмов (например, $\operatorname{tg}^2 x - \operatorname{tg} x - 1 = 0$).

Методы эти лучше называть *численными*, а не приближёнными, как принято у некоторых авторов, так как алгорифмы, к которым они сводятся, дают возможность получить значения корней с любой степенью точности, и в этом отношении они нисколько не уступают решениям, получаемым, например, по формулам, выражающим корни уравнения в радикалах.

Отметим, кстати, что вопрос о разрешимости уравнений в радикалах, важный в некоторых отношениях и для школьного преподавания (например, разрешимость задач на построение с помощью циркуля и линейки связана с возможностью представления корней некоторых уравнений с помощью квадратных радикалов), нас здесь не будет интересовать [1]).

---

[1]) Для ознакомления с этим вопросом отсылаем читателя к статье Л. Я. Окунева в этой книге.

ГЛАВА I

РЕШЕНИЕ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

§ 1. Постановка задачи

Рассмотрим сначала методы, пригодные в основном для численного решения алгебраических уравнений, хотя некоторые из них могут быть приспособлены и к решению уравнений трансцендентных.

Иногда при решении алгебраического уравнения нас интересует значение какого-нибудь одного корня (например, определяя сторону правильного восемнадцатиугольника, вписанного в круг единичного радиуса, мы убеждаемся в том, что она является меньшим положительным корнем уравнения $x^3 - 3x + 1 = 0$). В других же случаях надо вычислить значения всех действительных, а иногда и мнимых корней данного уравнения (например, при разложении рациональной дроби на сумму простейших дробей, при решении так называемых характеристических уравнений, соответствующих обыкновенным линейным дифференциальным уравнениям с постоянными коэффициентами и во многих других случаях).

Если нас интересуют все действительные (а может быть, и мнимые) корни данного уравнения, то полезно уметь быстро определять тот, по возможности узкий, интервал (или, соответственно, ту область на плоскости комплексного переменного), вне которого не могут лежать корни данного уравнения.

Как увидим несколько ниже, задача эта решается очень просто даже при отсутствии какого-либо представления о примерной величине корней уравнения.

Помимо установления области, в которой следует искать корни уравнения, для некоторых методов вычисления действительных корней [например, для способа Горнера (§. 4) или для способа Лагранжа (§ 5)] существенно предварительное отделение искомого корня от других корней, т. е. определение такого интервала, внутри которого лежал бы только один корень данного уравнения.

Эта задача, называемая *задачей отделения корней*, может быть также решена без предварительного представления о величине отдельных корней уравнения. Следующие два параграфа будут посвящены рассмотрению этих вопросов.

§ 2. Определение границ действительных корней

Существует ряд простых способов для быстрого определения верхней границы положительных корней уравнения с действительными коэффициентами:

$$f(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\ldots+a_{n-1}x+a_n=0, \qquad (1)$$

т. е. для определения такого положительного числа K, которое было бы заведомо больше любого корня данного уравнения.

Например, нетрудно убедиться, что верна следующая

Теорема. *Если* $a_0>0$ *и* A *— наибольшее из чисел*

$$|a_1|,\ |a_2|,\ \ldots,\ |a_n|,$$

то

$$K=1+\frac{A}{a_0}. \qquad (2)$$

Действительно, при любом положительном значении x мы имеем:

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_n\geqslant$$

$$\geqslant a_0x^n-Ax^{n-1}-Ax^{n-2}-\ldots-Ax-A \qquad (3)$$

(читателю полезно подумать, в каком случае здесь возможен знак равенства); отсюда

$$f(x)\geqslant a_0x^n-A\frac{x^n-1}{x-1}=x^n\left(a_0-\frac{A}{x-1}\right)+\frac{A}{x-1}. \qquad (3')$$

Если $x>1+\frac{A}{a_0}$, то $x-1>\frac{A}{a_0}>0$ или $a_0>\frac{A}{x-1}>0$. Следовательно, при $x>1+\frac{A}{a_0}$ оба члена справа в соотношении (3') будут положительными, а потому и подавно $f(x)>0$. Иначе говоря, ни при каком значении x, превышающем $1+\frac{A}{a_0}$, $f(x)$ не может обратиться в нуль, т. е. можно считать, что

$$K=1+\frac{A}{a_0}.$$

Примечание. С помощью совершенно аналогичных рассуждений можно было бы убедиться в том, что для уравнения с комплексными коэффициентами

$$b_0x^n+b_1x^{n-1}+\ldots+b_n=0$$

верхняя граница модулей корней уравнения равна $1+\frac{B}{|b_0|}$, где B — наибольший из модулей коэффициентов $|b_1|,\ |b_2|,\ \ldots,\ |b_n|$, т. е. мнимые корни этого уравнения следует искать среди тех точек

плоскости комплексного переменного, которые лежат внутри круга радиуса $R = 1 + \frac{B}{|b_0|}$ [1]).

Формула (2) даёт, вообще говоря, слишком крупные значения для K, так как при выводе её мы искали значения x, при которых положительной будет правая часть соотношения (3); последняя же может оказаться значительно меньшей, чем функция $f(x)$.

Лучшие значения для K даёт, как правило, способ Маклорена, в котором учитываются знаки коэффициентов уравнения (1); способ этот основан на следующей теореме:

Теорема. *Если в уравнении* (1)

$$a_0 > 0;\ a_1,\ a_2,\ \ldots,\ a_{m-1} \geqslant 0; \qquad a_m < 0 \tag{4}$$

и A_1 — *наибольшее из абсолютных значений отрицательных коэффициентов уравнения, то*

$$K = 1 + \sqrt[m]{\frac{A_1}{a_0}}. \tag{5}$$

Доказательство проводится так же, как и доказательство предыдущей теоремы. Именно при условиях (4) для любого положительного значения x

$$f(x) \geqslant a_0 x^n - A_1 (x^{n-m} + x^{n-m-1} + \ldots + x + 1) =$$

$$= a_0 x^n - \frac{A_1 (x^{n-m+1} - 1)}{x-1} = x^{n-m+1} \frac{a_0 x^{m-1}(x-1) - A_1}{x-1} + \frac{A_1}{x-1} >$$

$$> x^{n-m+1} \frac{a_0 (x-1)^m - A_1}{x-1} + \frac{A_1}{x-1}. \tag{6}$$

Но при $x > 1 + \sqrt[m]{\frac{A_1}{a_0}}$ имеем: $a_0 (x-1)^m > A_1$, в силу чего оба слагаемых в конце соотношений (6), а следовательно, и $f(x)$ будут положительны, т. е. корни уравнения (1) не могут превышать числá, определяемого равенством (5).

Укажем ещё на способ Ньютона, в основе которого лежит следующая

Теорема. *Если при каком-то значении* $x = c > 0$ *числа*

$$f(c),\ f'(c),\ f''(c),\ \ldots,\ f^{(n-1)}(c),\ f^{(n)}(c) > 0, \tag{7}$$

то можно считать, что $K = c$.

Действительно, из формулы Тейлора для многочленов [2])

$$f(x) \equiv f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \ldots + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

[1]) См., например, А. К. Сушкевич, Основы высшей алгебры, 1941, стр. 115.

[2]) См. стр. 321.

следует, что при $x \geqslant c$ и при условиях (7) $f(x)$ будет больше нуля, т. е. такие значения x не могут быть корнями уравнения (1).

Перейдём теперь к рассмотрению нижней границы отрицательных корней, т. е. такого числа, которое наверное меньше, чем любой отрицательный корень данного уравнения. Для её нахождения достаточно сделать в данном уравнении (1) подстановку

$$x = -z, \tag{8}$$

что приведёт к уравнению $f(-z) = 0$ или к уравнению

$$(-1)^n \cdot f(-z) = a_0 z^n - a_1 z^{n-1} + a_2 z^{n-2} - \ldots + (-1)^n \cdot a_n = 0 \tag{9}$$

(обе части уравнения умножены на $(-1)^n$ для сохранения знака старшего коэффициента).

Если для уравнения (9) найдём верхнюю границу положительных корней K_1, то для любого положительного корня этого уравнения будем иметь:

$$z_0 < K_1,$$

а в силу равенства (8) для соответствующего отрицательного корня уравнения (1) получим неравенство

$$x_0 = -z_0 > -K_1,$$

т. е. число $-K_1$ будет нижней границей отрицательных корней уравнения (1).

Поясним на примерах всё вышеизложенное (см. стр. 322—324).

Предварительно заметим, что при делении многочлена на двучлен вида $x - c$ вычисления удобнее всего располагать по так называемой схеме Горнера:

$$\begin{array}{c|cccccc} & a_0 & a_1 & a_2 & \ldots & a_{n-1} & a_n \\ \hline c & b_0 & b_1 & b_2 & \ldots & b_{n-1} & r \end{array}$$

$a_0\ a_1\ a_2\ \ldots\ a_{n-1}\ a_n$ ← коэффициенты делимого

$b_0\ b_1\ b_2\ \ldots\ b_{n-1}\ r$ ← остаток

$b_0\ b_1\ b_2\ \ldots\ b_{n-1}$ — коэффициенты частного

Очевидно, старшие коэффициенты делимого и частного всегда будут одинаковы ($b_0 = a_0$); любой же следующий коэффициент частного и остаток r вычисляются по формулам

$$\left.\begin{aligned} b_k &= a_k + b_{k-1} c, \\ r &= a_n + b_{n-1} c. \end{aligned}\right\} \tag{10}$$

Эти формулы сразу получаются из тождества

$$a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n \equiv$$
$$\equiv (x - c)(b_0 x^{n-1} + b_1 x^{n-2} + \ldots + b_{n-1}) + r,$$

если, раскрыв справа скобки и сделав приведение подобных членов, приравнять друг к другу коэффициенты при одинаковых степенях x в обеих частях тождества.

Схема Горнера особенно удобна при решении следующей задачи, которая в дальнейшем нам часто будет встречаться:

данный многочлен разложить по степеням двучлена $x-c$.

Пусть

$$\left.\begin{aligned} f(x) &= q_1(x)(x-c)+r_0,\\ q_1(x) &= q_2(x)(x-c)+r_1,\\ &\cdots\cdots\cdots\cdots\cdots\\ q_{n-2}(x) &= q_{n-1}(x)(x-c)+r_{n-2},\\ q_{n-1}(x) &= a_0(x-c)+r_{n-1}, \end{aligned}\right\} \qquad (11)$$

где $q_{s+1}(x)$ и r_s обозначают частное и остаток, получающийся при делении $q_s(x)$ на $x-c$.

Если самое последнее выражение в (11) подставить вместо $q_{n-1}(x)$ в предпоследнее равенство, а затем то, что при этом получится, подставить вместо $q_{n-2}(x)$ в третье равенство снизу и т. д., то придём, наконец, к тождеству

$$f(x)=r_0+r_1(x-c)+ \\ +r_2(x-c)^2+\ldots+r_{n-1}(x-c)^{n-1}+a_0(x-c)^n. \qquad (12)$$

Это и есть разложение данного многочлена по степеням $x-c$.

Если в тождестве (12) и в соотношениях, которые получатся после k-кратного его дифференцирования (в частности, при $k=n$), положим $x=c$, то получим:

$$\left.\begin{aligned} f(c) &= r_0,\\ f^{(k)}(c) &= k!\,r_k,\\ f^{(n)}(c) &= n!\,a_0. \end{aligned}\right\} \qquad (13)$$

Поэтому равенство (12) можно переписать в виде

$$f(x)=f(c)+f'(c)(x-c)+ \\ +\frac{f''(c)}{2!}(x-c)^2+\ldots+\frac{f^{(n-1)}(c)}{(n-1)!}(x-c)^{n-1}+\frac{f^{(n)}(c)}{n!}(x-c)^n, \qquad (14)$$

а это и есть формула Тейлора для многочленов (см. стр. 319).

Так как деление по схеме Горнера $f(x)$ на $x-c$ даёт коэффициенты частного $q_1(x)$, которое в свою очередь надо делить на $x-c$ и т. д., то все вычисления удобно располагать в одну таблицу, как это указано в следующем примере.

Разложить $f(x)=x^4-3x^3+x-2$ по степеням $x-3$:

	1	−3	0	1	−2		
3	1	0	0	1	1	$q_1(x)=x^3+1,$	$r_0=1;$
3	1	3	9	28		$q_2(x)=x^2+3x+9,$	$r_1=28;$
3	1	6	27			$q_3(x)=x+6,$	$r_2=27;$
3	1	9				$q_4(x)=1=a_0,$	$r_3=9.$

Следовательно,

$$f(x)\equiv 1\,(x-3)^4+9\,(x-3)^3+27\,(x-3)^2+28\,(x-3)+1;$$

при этом в силу соотношений (13)

$$f(3)=1,\ f'(3)=28,\ f''(3)=54,\ f'''(3)=54,\ f^{IV}(3)=24. \quad (15)$$

Обратимся к примерам вычисления границ корней.

Пример 1. Найти границы действительных корней уравнения

$$f(x)=2x^5+3x^4-10x^3-60x^2+40x+200=0.$$

Здесь $a_0=2$, $A=200$, $A_1=60$, $m=2$.

По первому способу $K=1+\frac{200}{2}=101$.

По способу Маклорена $K=1+\sqrt{\frac{60}{2}}\approx 6{,}5$ (приближённое значение берём с некоторым избытком).

Для определения числа c, удовлетворяющего условиям (7), приходится иногда делать несколько попыток; удобнее всего каждый раз применять схему Горнера. Попробуем испытать число 2:

	2	3	−10	−60	40	200
2	2	7	4	−52	−64	72
2	2	11	26	0	$-64=f'(2)$	

Так как $f'(2)=-64<0$, то попытка оказалась неудачной вследствие нарушения условий (7). Испытаем число 3:

	2	3	−10	−60	40	200
3	2	9	17	−9	13	239
3	2	15	62	177	544	

Так как вторая строка состоит исключительно из положительных чисел, то отпадает необходимость в вычислениях дальнейших строк, которые будут состоять из положительных чисел; в частности, числа $f(3)$, $f'(3)$, $f''(3)$, $f'''(3)$, $f^{IV}(3)$, $f^{V}(3)$ будут положительны. Итак, по способу Ньютона $K=3$. Таким образом, наилучшая верхняя граница положительных корней получилась по способу Ньютона.

Подстановка $x=-z$ приводит нас к уравнению

$$2z^5-3z^4-10z^3+60z^2+40z-200=0.$$

Здесь $a_0=2$, $A=200$, $A_1=200$, $m=1$. Следовательно, имеем:

По первому способу $K_1=1+\frac{200}{2}=101$.

По способу Маклорена $K_1=1+\frac{200}{2}=101$.

По способу Ньютона: испытаем число 2:

	2	−3	−10	60	40	−200
2	2	1	−8	44	128	56
2	2	5	2	+	+	

Так как получилась строчка исключительно положительных чисел, то условия (7) выполняются и $K_1=2$.

Опять способ Ньютона дал наилучший результат, в то время как первые два способа дали совершенно одинаковые и очень плохие значения для K_1.

Итак, нижняя граница отрицательных корней равна $-K_1=-2$; следовательно, все действительные корни уравнения (15) лежат в интервале (—2, 3).

Пример 2. Найти границы действительных корней уравнения

$$2x^6+10x^4-160x^2+x-12=0. \tag{16}$$

Отметим, что для данного уравнения $m=4$, так как перед первым отрицательным коэффициентом стоит четыре члена, коэффициенты которых неотрицательны (обязательно надо учитывать отсутствующие члены!).

Преобразованное уравнение имеет вид

$$2z^6+10z^4-160z^2-z-12=0. \tag{17}$$

Легко видеть, что по первому способу $K=81$, $K_1=81$, т. е. можно утверждать, что все действительные корни данного уравнения (16) (если они у него имеются) лежат в интервале $(-K_1, K)$, т. е. в интервале (—81, 81).

По способу Маклорена

$$K = K_1 = 1 + \sqrt[4]{\frac{160}{2}} \approx 4,$$

следовательно, можно утверждать, что корни лежат в интервале (—4, 4).

Переходя к способу Ньютона, испытаем число 3 сразу для уравнений (16) и (17) — они отличаются только знаком коэффициента при неизвестном в первой степени:

	2	0	10	0	—160	±1	—12
3	2	6	28	84	92	+	+

Так как вся строчка состоит только из положительных чисел, то 3 удовлетворяет условиям (7), а потому $K = 3$ и $K_1 = 3$; интервал же для действительных корней будет (—3, 3).

Упомянем ещё о двух границах корней, играющих значительно меньшую роль в вопросах решения уравнений: о *нижней границе положительных корней* (так называется положительное число, которое наверное меньше любого положительного корня данного уравнения), равной $\frac{1}{K_2}$, где K_2 — верхняя граница положительных корней уравнения, получающегося из данного подстановкой $x = \frac{1}{u}$, и о *верхней границе отрицательных корней* (так называется отрицательное число, которое наверное больше любого отрицательного корня данного уравнения), равной $-\frac{1}{K_3}$, где K_3 — верхняя граница положительных корней уравнения, получающегося из данного подстановкой $x = -\frac{1}{v}$.

§ 3. Отделение корней

Наиболее совершенное средство для отделения корней даёт теорема Штурма, к рассмотрению которой мы и перейдём.

Будем предполагать, что в предложенном нами уравнении с действительными коэффициентами кратные корни уже выделены, т. е. мы имеем дело с уравнением

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n = 0, \qquad (1)$$

все корни которого простые.

Составим систему функций

$$f(x),\ f'(x),\ R_1(x),\ R_2(x),\ \ldots,\ R_{m-1}(x),\ R_m = \text{const} \qquad (2)$$

следующим образом.

Деля $f(x)$ на $f'(x)$, обозначим получающийся при делении остаток через $-R_1(x)$; взятый же с обратным знаком, остаток этот даёт третью функцию $R_1(x)$ системы (2). Деля затем $f'(x)$ на $R_1(x)$, обозначим остаток через $-R_2(x)$, а взяв его с обратным знаком, получим четвёртую функцию системы (2). Это продолжим до тех пор, пока не получится функция, на которую предшествующая функция системы (2) делится без остатка.

Описанный процесс совпадает по существу с отысканием наибольшего общего делителя функций $f(x)$ и $f'(x)$ с помощью алгорифма Евклида [1]) (изменение знаков остатков не имеет значения, так как в алгорифме Евклида допускается умножение на постоянные числа не только окончательных остатков, но даже остатков промежуточных, получаемых в процессе деления). Так как уравнение (1) не имеет кратных корней, то $f(x)$ и $f'(x)$ будут взаимно простыми, следовательно, наибольший общий делитель их, т. е. последний остаток в описанном выше процессе, будет постоянным числом. Предположив, что остаток этот получается при m-м делении, обозначим его через $-R_m$; взяв же это число с обратным знаком, получим последний член системы (2).

Составленные таким образом функции (2) носят название *функций Штурма*, а вся система (2) называется *системой Штурма* уравнения (1).

Будем это название сохранять и для системы функций, отличающихся от соответствующих функций системы (2) постоянными положительными множителями.

Пусть при каком-то значении аргумента $x=c$ вычислены значения функций Штурма, т. е. получена система постоянных чисел

$$f(c),\ f'(c),\ R_1(c),\ \ldots,\ R_{m-1}(c),\ R_m. \qquad (3)$$

Если какие-нибудь два соседних числа этой системы имеют разные знаки, то будем говорить, что переход от одного числа к другому сопровождается переменой знака (при одинаковых же знаках у двух соседних чисел будем говорить, что имеет место повторение знака); если какое-нибудь из чисел ряда (3), кроме $f(c)$, будет равно нулю, то этот член будем пропускать, сравнивая друг с другом знаки соседних с ним чисел.

Обозначим общее число перемен знака в последовательности (3) через $W(c)$.

Теорема Штурма. *Если* $a<b$, *то* $W(a)\geqslant W(b)$, *причём разность* $W(a)-W(b)$ *равна числу действительных корней уравнения* (1), *заключённых в интервале* (a, b).

Другими словами, число перемен знака, теряемых системой функций Штурма при переходе от a к b, равно числу действительных корней уравнения (1), лежащих в интервале (a, b).

---

[1]) См. Э. э. м., кн. 1, А. Я. Хинчин, Основы теории чисел.

Поясним сначала теорему Штурма примером, а затем дадим её доказательство.

Пример. Пусть дано уравнение

$$x^3 - 10x + 2 = 0, \tag{4}$$

и надо, во-первых, найти число положительных и число отрицательных корней этого уравнения, а во-вторых, отделить корни друг от друга.

Так как $f'(x) = 3x^2 - 10$, то выполняем следующее деление:

$$\begin{array}{r|l} 3(x^3 - 10x + 2) & 3x^2 - 10 \\ \hline 3x^3 - 10x & x \\ \hline -20x + 6 = -R_1(x). & \end{array}$$

Следовательно, можно считать $R_1(x) = 10x - 3$.

Примечание. Мы умножили делимое на положительное число 3, чтобы не иметь дела с дробными коэффициентами; при этом остаток возрос в три раза, что допустимо, так как нас интересуют не сами числа $R_1(a)$ и $R_1(b)$, а лишь знаки этих чисел. По той же причине мы взяли $R_1(x) = 10x - 3$, а не $20x - 6$. Ясно, что недопустимым является умножение или сокращение на отрицательные числа.

Разделим теперь $f'(x)$ на $R_1(x)$:

$$\begin{array}{r|l} 10(3x^2 - 10) & 10x - 3 \\ \hline 30x^2 - 9x & 3x + 9 \\ \hline 10(9x - 100) & \\ \hline 90x - 27 & \\ \hline -973 = -R_2. & \end{array}$$

Следовательно, $R_2 = 973$, или можно также считать, что $R_2 = 1$.

Итак, мы получили систему функций Штурма для уравнения (4):

$$f(x) = x^3 - 10x + 2; \quad f'(x) = 3x^2 - 10; \quad R_1(x) = 10x - 3; \quad R_2 = 1.$$

Имея в виду определить число перемен знака $W(x)$ при разных значениях x, будем заносить значения этих функций в следующую таблицу:

x	Значения или знаки функций				$W(x)$
	$f(x)$	$f'(x)$	$R_1(x)$	R_2	
0	+2	−10	− 3	+1	2
$+\infty$	+	+	+	+1	0
$-\infty$	−	+	−	+1	3
− 2	+14	+2	−23	+1	2
− 4	−22	+38	−43	+1	3
5	+	+	+	+1	0
2	−	+	+	+1	1

Желая найти общее число положительных корней уравнения, прежде всего находим $W(0)=2$, а затем полагаем, условно выражаясь, $x=+\infty$.

Конечно, значений функций Штурма при $x=+\infty$ мы не вычисляем, так как $+\infty$ есть лишь символ переменной величины, принимающей какие угодно большие положительные значения, но при определении знаков $f(x)$, $f'(x)$ и т. д. при $x=+\infty$ мы руководствуемся простым правилом, в силу которого знак любого многочлена при достаточно больших по абсолютной величине значениях аргумента совпадает со знаком его старшего члена (это следует из того, что старший член при $|x|\to+\infty$ растёт по абсолютной величине значительно быстрее, чем остальные члены многочлена [1])).

Итак, число положительных корней равно

$$W(0)-W(+\infty)=2-0=2.$$

Найдя затем $W(-\infty)=3$, мы можем утверждать, что число отрицательных корней уравнения равно

$$W(-\infty)-W(0)=1.$$

Желая найти сравнительно небольшой интервал, заключающий отрицательный корень уравнения, полагаем $x=-2$, и так как

$$W(-2)=2=W(0),\quad \text{а}\quad W(-\infty)-W(-2)=1,$$

то единственный отрицательный корень будет меньше чем -2; но $W(-4)=3$; следовательно, можно утверждать, что отрицательный корень лежит в интервале $(-4, -2)$.

Положив $x=5$, находим: $W(5)=0$, т. е. оба положительных корня лежат в интервале $(0, 5)$, так как

$$W(0)-W(5)=2.$$

Наконец, при $x=2$ имеем: $W(2)=1$, следовательно, и положительные корни отделены друг от друга: один лежит в интервале $(0, 2)$, а другой — в интервале $(2, 5)$, так как

$$W(0)-W(2)=W(2)-W(5)=1.$$

Перейдём к доказательству теоремы Штурма; оно основано на следующих четырёх свойствах функций Штурма:

I. *При изменении x последняя функция Штурма не меняет своего знака.*

Это свойство очевидно, так как последняя функция есть просто постоянное число.

---

[1]) См., например, А. К. Сушкевич, Основы высшей алгебры, 1941, стр. 131.

II. *Две соседние функции Штурма не могут обратиться в нуль при одном и том же значении аргумента.*

Для доказательства возьмём ряд тождеств (вытекающих из известного соотношения: делимое равно делителю, умноженному на частное, плюс остаток)

$$\left.\begin{aligned}
f(x) &\equiv f'(x)\ q_1(x) - R_1(x),\\
f'(x) &\equiv R_1(x)\ q_2(x) - R_2(x),\\
R_1(x) &\equiv R_2(x)\ q_3(x) - R_3(x),\\
&\cdots\cdots\cdots\cdots\cdots\\
R_{k-1}(x) &\equiv R_k(x)\ q_{k+1}(x) - R_{k+1}(x),\\
&\cdots\cdots\cdots\cdots\cdots\\
R_{m-2}(x) &\equiv R_{m-1}(x)\ q_m(x) - R_m,\\
R_{m-1}(x) &\equiv R_m\ q_{m+1}(x).
\end{aligned}\right\} \qquad (5)$$

(Здесь через $q_1(x)$, $q_2(x)$ и т. д. обозначены соответствующие частные.) Если бы при каком-то $x = x_0$ какие-нибудь две соседние функции равнялись нулю, например $R_k(x_0) = R_{k+1}(x_0) = 0$, то из соответствующего соотношения в (5) следовало бы, что $R_{k-1}(x_0) = 0$, а из предшествующих ему соотношений последовательно вытекало бы, что $R_{k-2}(x_0) = 0$, $R_{k-3}(x_0) = 0, \ldots$, $R_2(x_0) = 0$, $R_1(x_0) = 0$, и, наконец, $f'(x_0) = 0$ и $f(x_0) = 0$.

Но это означало бы, что x_0 есть кратный корень уравнения (1), а это противоречит условию, в силу которого все корни уравнения (1) простые.

III. *Если какая-нибудь из функций Штурма* $f'(x)$, $R_1(x)$, $R_2(x), \ldots, R_{m-1}(x)$ *обращается в нуль при некотором значении* $x = c$, *то значения соседних с нею функций Штурма, вычисленные при том же значении* $x = c$, *имеют разные знаки.*

Действительно, если $R_k(c) = 0$, то из тождества

$$R_{k-1}(x) \equiv R_k(x)\ q_{k+1}(x) - R_{k+1}(x)$$

следует:

$$R_{k-1}(c) = -R_{k+1}(c). \qquad (6)$$

Свойство III играет существенную роль в доказательстве теоремы Штурма благодаря вытекающему из него следствию:

Пусть x принимает настолько близкие друг к другу значения $c - \varepsilon$, c, $c + \varepsilon$, что знаки чисел $R_{k-1}(c - \varepsilon)$ и $R_{k-1}(c + \varepsilon)$ совпадают со знаком числа $R_{k-1}(c)$, а знаки чисел $R_{k+1}(c - \varepsilon)$ и $R_{k+1}(c + \varepsilon)$ — со знаком числа $R_{k+1}(c)$ [1]). Тогда в силу соотношения (6) в

[1]) Это выполнимо, так как $R_{k-1}(x)$ и $R_{k+1}(x)$, будучи многочленами, являются функциями непрерывными.

каждой тройке чисел

$$R_{k-1}(c-\varepsilon),\quad R_k(c-\varepsilon),\quad R_{k+1}(c-\varepsilon), \tag{7}$$

$$R_{k-1}(c+\varepsilon),\quad R_k(c+\varepsilon),\quad R_{k+1}(c+\varepsilon) \tag{8}$$

крайние числа будут иметь разные знаки, и независимо от того, какие знаки будут у чисел $R_k(c-\varepsilon)$ и $R_k(c+\varepsilon)$, каждая из этих троек чисел будет давать одну перемену знака.

Очевидно, всё изложенное выше относится и к тройке функций $f(x)$, $f'(x)$, $R_1(x)$ при условии, что x, возрастая, проходит через корень функции $f'(x)$. С другой стороны, те из функций Штурма, которые в точке $x=c$ не обращаются в нуль, сохраняют свой знак в малой окрестности этой точки, а потому никакого влияния на изменение числа перемен знака оказать не могут. Таким образом, из свойства III вытекает следствие:

Если x, возрастая, проходит через корень одной из функций $f'(x)$, $R_1(x)$, $R_2(x)$, ..., $R_{m-1}(x)$, то при этом число перемен знака среди функций полной системы Штурма (3) *не меняется.*

Иначе обстоит дело, как это следует из свойства IV, когда x, возрастая, проходит через корень функции $f(x)$:

IV. *Если x_1 — корень функции $f(x)$, то при достаточно малом положительном ε числа $f(x_1-\varepsilon)$ и $f'(x_1-\varepsilon)$ имеют разные, а числа $f(x_1+\varepsilon)$ и $f'(x_1+\varepsilon)$ — одинаковые знаки.*

Действительно, так как $f'(x_1)\neq 0$ (иначе x_1 было бы кратным корнем $f(x)$), то, взяв ε настолько малым, чтобы в интервале $(x_1-\varepsilon,\ x_1+\varepsilon)$ функция $f(x)$ сохраняла знак, будем иметь одну из двух возможностей:

1) В интервале $(x_1-\varepsilon,\ x_1+\varepsilon)$ функция $f'(x)>0$. Но тогда $f(x)$ есть функция, возрастающая в этом интервале, и следовательно, $f(x_1-\varepsilon)<0$ и $f(x_1+\varepsilon)>0$, в то время как $f'(x_1-\varepsilon)>0$ и $f'(x_1+\varepsilon)>0$.

2) В интервале $(x_1-\varepsilon,\ x_1+\varepsilon)$ функция $f'(x)<0$. Но тогда $f(x)$ есть функция, убывающая в этом интервале, и следовательно, $f(x_1-\varepsilon)>0$ и $f(x_1+\varepsilon)<0$, в то время как $f'(x_1-\varepsilon)<0$ и $f'(x_1+\varepsilon)<0$. Таким образом, свойство IV доказано полностью.

В силу этого свойства

$$W(x_1-\varepsilon)-W(x_1+\varepsilon)=1,$$

т. е. система функций Штурма теряет одну перемену знака всякий раз, когда x, возрастая, проходит через корень уравнения (1).

Так как во всех других случаях число перемен знака в системе функций Штурма не меняется, то общее число теряемых перемен знака при переходе от a к b будет равно числу корней уравнения (1), лежащих в интервале $(a,\ b)$, а это и есть теорема Штурма.

Как показывает рассмотренный выше пример, с помощью теоремы Штурма действительные корни могут быть отделены друг от

друга, и каждый интервал, заключающий тот или иной корень, может быть затем уменьшен до требуемых размеров.

Однако, установив, что в каком-то интервале лежит точно один корень данного уравнения, для дальнейшего уменьшения интервала нет необходимости вычислять значения всех функций Штурма, а достаточно определять всё более и более узкие интервалы, на концах которых функция $f(x)$ имеет разные знаки (табличный способ вычисления корней).

Казалось бы, можно было бы с самого начала, не прибегая к теореме Штурма, табличным путём отделить корни друг от друга; во многих случаях так и делают. Однако при этом существует опасность не заметить корней уравнения, лежащих в каком-нибудь интервале. Например, для уравнения

$$F(x) = 9x^3 - 61x + 60 = 0,$$

составляя таблицу

x	$F(x)$
0	60
1	8
2	10
3	120

можно было бы сделать заключение, что уравнение не имеет положительных корней (так как при $x > 3$ функция $F(x)$ будет принимать лишь положительные значения, ибо старший член её будет расти гораздо быстрее, чем отрицательный член).

Однако такое заключение было бы неверным, так как в интервале (1, 2) лежат два действительных корня, в существовании которых мы убеждаемся, определив, например, $F\left(\frac{3}{2}\right) = -\frac{9}{8}$.

Правда, этот пример подсказывает следующий практический совет: когда убывание функции сменяется её возрастанием (или наоборот), а сама функция, принимая сравнительно малые значения, не меняет знака при табличных значениях аргумента, надо быть осторожным и лишь после детальной проверки поведения функции в соответствующем интервале утверждать, что в этом интервале нет корней уравнения.

Еще опаснее случай, когда при наличии в каком-нибудь единичном интервале трёх действительных корней таблица «подсказывает», что этот интервал содержит лишь один действительный корень. Так, например, будет обстоять дело для уравнения

$$\Phi(x) = (3x - 7)(4x - 11)(x^2 - 3x + 1) =$$
$$= 12x^4 - 97x^3 + 272x^2 - 292x + 77 = 0.$$

Из таблицы

x	$\Phi(x)$
0	77
1	−28
2	− 3
3	+ 2
4	125

можно было бы заключить, что в интервале (2, 3) заключён один корень уравнения; на самом же деле в этом интервале заключено три корня:

$$x_1 = \frac{7}{3}; \quad x_2 = \frac{11}{4}; \quad x_3 = \frac{3}{2} + \sqrt{1{,}25} \approx 2{,}6.$$

Итак, табличный способ может привести к ошибочному заключению, тогда как способ Штурма в этом отношении безупречен, хотя и связан иногда с очень громоздкими вычислениями.

Кроме способа Штурма определения числа корней, заключённых в данном интервале, существуют другие приёмы. Приведём без доказательства теорему Бюдана-Фурье, связанную с гораздо более простой системой функций, чем система Штурма, но иногда дающую, к сожалению, довольно расплывчатый ответ.

Теорема Бюдана-Фурье. *Число действительных корней уравнения* (1), *заключённых в интервале* (a, b), *равно числу теряемых перемен знака в системе функций*

$$f(x),\ f'(x),\ f''(x),\ \ldots,\ f^{(n-1)}(x),\ f^{(n)}(x)$$

при переходе от a *к* b *или на чётное число меньше, чем это число.*

Если, например, теряется пять перемен знака, то по теореме Бюдана-Фурье можно утверждать, что число корней, лежащих в данном интервале, равно 5, 3 или 1.

Правда, если теряется одна перемена знака, то можно утверждать, что в интервале (a, b) лежит точно один корень уравнения.

Без доказательства же приведём теорему Декарта [1]).

Теорема Декарта. *Число положительных корней уравнения* (1) *равно числу перемен знака в последовательности коэффициентов уравнения* $a_0, a_1, a_2, \ldots, a_{n-1}, a_n$ *или на чётное число меньше, чем это число.*

Теорема эта позволяет быстро (но не всегда точно) определять число положительных корней уравнения, а при использовании подстановки $x = -z$ — число его отрицательных корней.

---

[1]) Интересующихся подробностями, связанными с теоремами Бюдана-Фурье и Декарта, отошлём к цитированной выше книге Сушкевича.

§ 4. Способ Горнера

Пусть $$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0 \qquad (1)$$

— уравнение с действительными коэффициентами.

Для вычисления какого-нибудь действительного корня уравнения (1) способом Горнера поступают следующим образом: определив хотя бы табличным способом целую часть искомого корня, с помощью некоторых подстановок переходят затем к такому уравнению, у которого целая часть соответствующего корня равна первому десятичному знаку искомого корня. Повторяя этот процесс, можно получить сколько угодно десятичных знаков интересующего нас корня уравнения (1).

Каковы же эти подстановки?

Пусть $x_1=c_0+\frac{c_1}{10}+\frac{c_2}{100}+\frac{c_3}{1000}+\ldots=c_0,c_1c_2c_3\ldots$ представляет искомый корень, написанный в виде десятичной дроби, и мы уже нашли c_0 табличным способом, установив, например, что числа $f(c_0)$ и $f(c_0+1)$ имеют разные знаки. Тогда подстановка

$$x-c_0=y \qquad (2)$$

приведёт нас к уравнению

$$\varphi_1(y)=0, \qquad (3)$$

у которого соответствующий корень $y_1=0,c_1c_2c_3\ldots$, а затем подстановка

$$y=\frac{Y}{10} \qquad (4)$$

заменит уравнение (3) уравнением

$$\psi_1(Y)=0, \qquad (5)$$

где $Y_1=c_1,c_2c_3c_4\ldots$

Определив целую часть корня Y_1, подстановкой

$$Y-c_1=z \qquad (6)$$

преобразуем уравнение (5) в уравнение

$$\varphi_2(z)=0, \qquad (7)$$

а последнее подстановкой

$$z=\frac{Z}{10} \qquad (8)$$

в уравнение

$$\psi_2(Z)=0, \qquad (9)$$

у которого соответствующий корень

$$Z_1=c_2,c_3c_4\ldots$$

Ясно, что процесс этот можно повторять сколько угодно раз.

Если бы мы нашли Z_1 не с точностью до единицы, а, например, с точностью до одной сотой, т. е. установили бы, что

$$c_2, c_3c_4 < Z_1 < c_2, c_3c_4 + 0{,}01,$$

то это нам дало бы сразу три цифры искомого корня, а следующие подстановки имели бы вид

$$Z - c_2, c_3c_4 = u, \tag{10}$$

$$u = \frac{U}{1000}. \tag{11}$$

Подстановки (2), (6), (10) и т. д. удобнее всего выполнять, предварительно разлагая $f(x)$ по степеням $x - c_0$, $\psi_1(Y)$ — по степеням $Y - c_1$ и т. п.; переход от уравнения (3) к уравнению (5) сопровождается лишь умножением коэффициентов соответственно на 1, 10, 10^2, 10^3 и т. д. (при подстановке же (11) — на 1, 10^3, 10^6 и т. д.).

Пример 1. Найдём, чему равен положительный корень уравнения

$$f(x) = x^3 + x - 13 = 0.$$

Так как $f(2) = -3$, а $f(3) = 17$, то $2 < x_1 < 3$, т. е. $c_0 = 2$. Следовательно, надо сделать подстановки $x - 2 = y$, $y = \frac{Y}{10}$ или, объединяя их: $x = 2 + \frac{Y}{10}$.

	1	0	1	−13
2	1	2	5	−3
2	1	4	13	
2	1	6		

Таким образом, имеем:

$$\varphi_1(y) = y^3 + 6y^2 + 13y - 3 = 0,$$

$$\psi_1(Y) = Y^3 + 60Y^2 + 1300Y - 3000 = 0.$$

Здесь сразу видно, что $\psi_1(2) < 0$, а $\psi_1(3) > 0$; следовательно, $2 < Y_1 < 3$, т. е. $c_1 = 2$, и надо сделать подстановки $Y - 2 = z$; $z = \frac{Z}{10}$ или, объединяя их: $Y = 2 + \frac{Z}{10}$ (для искомого же корня будем иметь: $x_1 = 2{,}2 + \frac{Z_1}{100}$):

	1	60	1300	−3000
2	1	62	1424	−152
2	1	64	1552	
2	1	66		

откуда

$$\psi_2(Z) = Z^3 + 660Z^2 + 155\,200Z - 152\,000 = 0.$$

Ясно, что $0 < Z_1 < 1$, и соответствующие подстановки будут: $Z = 0 + u = u$; $u = \frac{U}{10}$ или $Z = \frac{U}{10}$ (следовательно, $x_1 = 2{,}20 + \frac{U_1}{1000}$); это даст нам:

$$\psi_3(U) = U^3 + 6600U^2 + 15\,520\,000U - 152\,000\,000 = 0.$$

Хотя мы и получили уравнение с очень крупными коэффициентами, но резкое возрастание их к концу уравнения позволит нам получить границы для U_1 простым делением свободного члена на коэффициент при U, причём деление это позволит даже найти сразу несколько следующих цифр искомого корня:

$$\begin{array}{r|l} 152\,000\,000 & 15\,520\,000 \\ \underline{139\,680\,000} & \overline{9{,}75} \\ 12\,320\,000{,}0 & \\ \underline{10\,864\,000{,}0} & \\ 1\,456\,000{,}00 & \\ \underline{776\,000{,}00} & \\ 680\,000{,}00 & \end{array}$$

Последняя цифра частного 5 сотых (а не 9 сотых, как должно было бы получиться при обычном делении) взята с таким расчётом, чтобы получился остаток (равный 680 000), несколько бóльший, чем значение двух временно отброшенных членов уравнения $U^3 + 6\,600U^2$ при $U = 9{,}75$:

$$9{,}75^3 + 6600 \cdot 9{,}75^2 < 661\,000 < 680\,000.$$

Следовательно, $\psi_3(9{,}75) < 0$.

С другой стороны, если бы мы взяли последнюю цифру в частном равной 6 сотым, то получился бы остаток

$$524\,800 < 9{,}76^3 + 6600 \cdot 9{,}76^2,$$

и следовательно, мы имели бы $\psi(9{,}76) > 0$.

Итак, $9{,}75 < U_1 < 9{,}76$, откуда

$$2{,}20975 < x_1 < 2{,}20976,$$

т. е. корень уравнения получен с точностью до 10^{-5}. В силу теоремы Декарта других положительных корней данное уравнение не имеет.

Конечно, в интервале $(c_0,\ c_0 + 1)$ может лежать не один, а несколько корней уравнения (1); заметим, что при чётном числе корней знаки $f(c_0)$ и $f(c_0 + 1)$ одинаковы, и при неосторожном применении табличного способа можно, как уже указывалось в примечании к теореме Штурма (§ 3), такие корни даже проглядеть.

Однако, если будет установлено, что в интервале $(c_0,\ c_0 + 1)$ лежат, например, точно два корня, то каждый из них может быть

вычислен отдельно, как было указано в разобранном примере. В самом деле, пусть

$$c_0 < x_1 < c_0 + 1 \quad \text{и} \quad c_0 < x_2 < c_0 + 1.$$

Тогда уравнение (5) будет иметь корни

$$0 < Y_1 < 10 \quad \text{и} \quad 0 < Y_2 < 10.$$

Если окажется, что целые части у Y_1 и Y_2 совпадут, т. е. что Y_1 и Y_2 лежат между c_1 и $c_1 + 1$, то переход к уравнению (9) совершим с помощью подстановки $Y = c_1 + \frac{Z}{10}$.

Пусть целые части у Z_1 и Z_2 будут разные:

$$c_2' < Z_1 < c_2' + 1 \quad \text{и} \quad c_2'' < Z_2 < c_2'' + 1,$$

тогда подстановка $Z = c_2' + \frac{U}{10}$ приведёт к цепи вычислений, дающих x_1 с требуемой степенью точности; подстановка же $Z = c_2'' + \frac{V}{10}$ приведёт в конечном счёте к корню x_2.

Пример 2. Решим уравнение

$$9x^3 - 61x + 60 = 0.$$

Из § 3 мы знаем, что оно имеет в интервале (1, 2) два корня; следовательно, $1 < x_1, x_2 < 2$.

Сделаем подстановку $x = 1 + \frac{Y}{10}$.

	9	0	−61	60
1	9	9	−52	8
1	9	18	−34	
1	9	27		

$$\psi_1(Y) = 9Y^3 + 270Y^2 - 3400Y + 8000 = 0.$$

Из таблицы

Y	3	4	6	7
знак $\psi_1(Y)$	+	−	−	+

видно, что $3 < Y_1 < 4$, $6 < Y_2 < 7$.

Делая подстановку $Y = 3 + \frac{Z}{10}$, придём к уравнению

$$9Z^3 + 3510Z^2 - 153\,700Z + 473\,000 = 0$$

(читателю предоставляется самому получить это уравнение); так как $3 < Z_1 < 4$, то $1{,}33 < x_1 < 1{,}34$. Конечно, продолжая вычисления, можно получить сколько угодно десятичных знаков для x_1.

Подстановка же $Y=6+\frac{W}{10}$ приведёт нас к уравнению

$$9W^3+4320W^2+81\,200W-736\,000=0,$$

и так как $6<W_1<7$, то $1{,}66<x_2<1{,}67$.

Сделаем, наконец, замечание о вычислении по способу Горнера отрицательных корней.

Возьмём для примера уравнение

$$f(x)=x^3+10x+53=0.$$

Так как $f(-3)=-4$, а $f(-2)=25$, то $-3<x_1<-2$. Делая подстановку $x=-3+\frac{Y}{10}$:

	1	0	10	53
−3	1	−3	19	−4
−3	1	−6	37	
−3	1	−9		

придём к уравнению

$$Y^3-90Y^2+3700Y-4000=0,$$

где $1<Y_1<2$. Следовательно, надо сделать подстановку $Y=1+\frac{Z}{10}$, которая приведёт к уравнению

$$Z^3-930Z^2+351\,700Z-391\,000=0,$$

причём $1<Z_1<2$.

Таким образом, используя приём, применяемый для обозначения логарифмов с отрицательной характеристикой, получим:

$$\bar{3},11<x_1<\bar{3},12,$$

или

$$-2{,}89<x_1<-2{,}88.$$

К тому же результату можно было бы придти, сделав с самого начала в данном уравнении замену: $x=-w$, что привело бы к уравнению $w^3+10w-53=0$. Найдя для последнего $2{,}88<w_1<2{,}89$, мы получим: $-2{,}89<x_1=-w_1<-2{,}88$.

§ 5. Способ Лагранжа

В то время как в способе Горнера ищутся один за другим десятичные знаки действительного корня уравнения, в способе Лагранжа последовательно определяются так называемые неполные частные $q_0, q_1, q_2, \ldots$ той непрерывной дроби, в которую разлагается искомый корень уравнения.

Пусть для уравнения с действительными коэффициентами

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0 \tag{1}$$

установлено (например, табличным способом), что корень лежит в интервале $(q_0,\ q_0+1)$, т. е.

$$x_1=q_0+\alpha=q_0+\frac{1}{y}, \tag{2}$$

где α — неизвестная положительная величина, меньшая единицы, а y — неизвестная положительная величина, бóльшая единицы.

Ясно, что подстановка

$$x=q_0+\frac{1}{y} \tag{3}$$

приведёт нас к уравнению

$$F(y)=0, \tag{4}$$

у которого обязательно найдётся положительный корень, превышающий единицу [если в интервале $(q_0,\ q_0+1)$ заключено несколько корней уравнения (1), то столько же положительных, превышающих единицу корней будет иметь и уравнение (4)].

Обозначим корень уравнения (4), соответствующий корню x_1, через y_1; тогда будем иметь:

$$x_1=q_0+\frac{1}{y_1}. \tag{3'}$$

Если найдём затем такое целое число q_1, что

$$q_1<y_1<q_1+1,$$

то подстановка

$$y=q_1+\frac{1}{z} \tag{5}$$

приведёт нас к уравнению

$$\Phi(z)=0, \tag{6}$$

имеющему положительный, превышающий единицу корень z_1, причём

$$y_1=q_1+\frac{1}{z_1} \tag{5'}$$

(читателю полезно подумать, в каком случае уравнение (6) может иметь несколько положительных корней, превышающих единицу).

Найдя интервал $(q_2,\ q_2+1)$, в котором заключён z_1, делаем подстановку

$$z=q_2+\frac{1}{u} \tag{7}$$

и продолжаем этот процесс до тех пор, пока не будет обеспечена возможность определения искомого корня x_1 с заданной точностью.

Сопоставляя (3′), (5′) и т. п., получим представление искомого корня в виде непрерывной (или цепной) дроби

$$x_1 = q_0 + \frac{1}{y_1} = q_0 + \frac{1}{q_1 + \frac{1}{z_1}} = q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \frac{1}{u_1}}} = \ldots =$$

$$= q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ldots + \cfrac{1}{q_k + \cfrac{1}{w_1}}}}. \qquad (8)$$

Конечно, может случиться, что положительный корень одного из получающихся в результате преобразований уравнений будет числом целым; пусть, например, окажется, что $u_1 = q_3$ (целому числу); тогда

$$x_1 = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3}}}, \qquad (9)$$

т. е. искомый корень будет числом рациональным.

Таким образом, рациональность искомого корня обнаруживается в способе Лагранжа автоматически, что представляет принципиальное преимущество способа Лагранжа перед другими методами вычисления действительных корней. Для корней же иррациональных процесс никогда не закончится, т. е. x_1 будет выражаться бесконечной непрерывной дробью, для которой по способу Лагранжа можно вычислить сколько угодно неполных частных: q_0, q_1, q_2 и т. д.

Если непрерывную дробь (конечную или бесконечную) оборвать на каком-то звене, то получится её так называемая *подходящая дробь*

$$\frac{P_s}{Q_s} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ldots + \cfrac{1}{q_s}}}.$$

Например,

$$\frac{P_0}{Q_0} = q_0 = \frac{q_0}{1}; \quad \frac{P_1}{Q_1} = q_0 + \frac{1}{q_1} = \frac{q_0 q_1 + 1}{q_1};$$

$$\frac{P_2}{Q_2} = q_0 + \frac{1}{q_1 + \frac{1}{q_2}} = \frac{q_0 q_1 q_2 + q_0 + q_2}{q_1 q_2 + 1}.$$

Подходящие дроби данной непрерывной дроби обладают рядом замечательных свойств; отметим здесь некоторые из них без доказательства [1]).

[1]) См. Э. э. м., кн. 1, А. Я. Хинчин, Элементы теории чисел.

I. *Существуют соотношения, выражающие числитель* (или знаменатель) *любой подходящей дроби через числители* (соответственно, через знаменатели) *двух предшествующих дробей*:

$$P_{k+1} = P_k q_{k+1} + P_{k-1}, \tag{10}$$

$$Q_{k+1} = Q_k q_{k+1} + Q_{k-1}. \tag{11}$$

Формулы (10), (11) позволяют очень просто вычислять одну за другой подходящие дроби, если предварительно вычислены $\frac{P_0}{Q_0}$ и $\frac{P_1}{Q_1}$; возьмём для примера разложение в непрерывную дробь числа $\frac{859}{392}$:

$$\frac{859}{392} = 2 + \cfrac{1}{5 + \cfrac{1}{4 + \cfrac{1}{2 + \cfrac{1}{2 + \frac{1}{3}}}}}.$$

Так как

$$\frac{P_0}{Q_0} = 2 = \frac{2}{1} \quad \text{и} \quad \frac{P_1}{Q_1} = 2 + \frac{1}{5} = \frac{11}{5}, \quad \text{а} \quad q_2 = 4,$$

то

$$\left.\begin{aligned} P_2 &= P_1 q_2 + P_0 = 11 \cdot 4 + 2 = 46, \\ Q_2 &= Q_1 q_2 + Q_0 = 5 \cdot 4 + 1 = 21. \end{aligned}\right\} \tag{12}$$

Вычисления удобно расположить в таблицу:

k	0	1	2	3	4	5
P_k	2	11	46	103	252	859
Q_k	1	5	21	47	115	392
q_k	2	5	4	2	2	3

Заполнив с самого начала верхнюю и нижнюю строки, а также первые два столбца средних строк, вычисляем [см. (12)] P_2 и Q_2, затем $P_3 = 103$ (умножив $P_2 = 46$ на следующее неполное частное $q_3 = 2$ и прибавляя $P_1 = 11$), аналогично Q_3 и т. д. Очевидно, $\frac{P_5}{Q_5}$ должно совпасть со значением всей непрерывной дроби. Итак, для данного примера подходящие дроби будут:

$$\frac{2}{1}, \ \frac{11}{5}, \ \frac{46}{21}, \ \frac{103}{47}, \ \frac{252}{115}, \ \frac{859}{392}.$$

II. *Все подходящие дроби с чётными номерами*

$$\frac{P_0}{Q_0},\ \frac{P_2}{Q_2},\ \frac{P_4}{Q_4},\ \ldots$$

будут меньше, а все подходящие дроби с нечётными номерами

$$\frac{P_1}{Q_1},\ \frac{P_3}{Q_3},\ \frac{P_5}{Q_5},\ \ldots$$

будут больше, чем данная непрерывная дробь.

III. *Всегда*

$$\frac{P_{s+1}}{Q_{s+1}}-\frac{P_s}{Q_s}=\frac{(-1)^s}{Q_sQ_{s+1}}. \tag{13}$$

Например, в рассмотренном случае

$$\frac{P_2}{Q_2}-\frac{P_1}{Q_1}=\frac{46}{21}-\frac{11}{5}=-\frac{1}{105},$$
$$\frac{P_3}{Q_3}-\frac{P_2}{Q_2}=\frac{103}{47}-\frac{46}{21}=\frac{1}{47\cdot 21}.$$

Последнее свойство даёт возможность определять в способе Лагранжа, в какой момент можно прекратить вычисления. Действительно, так как x_1 заключено между $\frac{P_s}{Q_s}$ и $\frac{P_{s+1}}{Q_{s+1}}$, то

$$\left|x_1-\frac{P_s}{Q_s}\right|<\left|\frac{P_{s+1}}{Q_{s+1}}-\frac{P_s}{Q_s}\right|=\frac{1}{Q_sQ_{s+1}}. \tag{14}$$

Но

$$Q_{s+1}=Q_sq_{s+1}+Q_{s-1}\geqslant Q_s+Q_{s-1}\quad(\text{так как } q_{s+1}\geqslant 1);$$

поэтому

$$\left|x_1-\frac{P_s}{Q_s}\right|<\frac{1}{Q_s(Q_s+Q_{s-1})}.$$

Если погрешность приближённого равенства $x_1\approx\frac{P_k}{Q_k}$ должна быть меньше заданного числа α, т. е. $\left|x_1-\frac{P_k}{Q_k}\right|<\alpha$, то, как только будем иметь:

$$\frac{1}{Q_s(Q_s+Q_{s-1})}<\alpha\quad\text{или}\quad Q_s(Q_s+Q_{s-1})>\frac{1}{\alpha},$$

вычисления можно будет прекратить.

Таким образом, если одновременно с вычислением q_0, q_1, q_2 и т. д. по способу Лагранжа заполнять таблицу (аналогичную таблице, указанной в примере) и вычислять значения выражения $Q_s(Q_s+Q_{s-1})$, то в любой момент легко будет определить, с какой точностью соответствующая подходящая дробь выражает искомый корень. Свойство же II даст возможность определить: найдено ли приближённое значение корня с недостатком или с избытком.

Поясним способ Лагранжа примерами. Предварительно заметим, что подстановка $x=q_0+\frac{1}{y}$ (и ей подобные) очень просто выполняется по схеме Горнера: ищутся коэффициенты разложения функции $f(x)$ по степеням $x-q_0$ и так как $x-q_0=\frac{1}{y}$, то для получения уравнения с неизвестным y достаточно взять полученные по схеме Горнера коэффициенты в обратном порядке (свободный член сделать старшим коэффициентом и т. д.; старший коэффициент превратить в свободный член).

Пример 1. Найти положительный корень уравнения

$$f(x)=7x^3-3x^2+4x-20=0.$$

Из таблицы

x	0	1	2
$f(x)$	-20	-12	32

видно, что положительный корень уравнения лежит в интервале (1, 2), следовательно, надо сделать подстановку $x=1+\frac{1}{y}$:

	7	-3	4	-20
1	7	4	8	-12
1	7	11	19	
1	7	18		

Беря полученные числа в обратном порядке (и умножая их на -1, чтобы старший коэффициент преобразованного уравнения сделался положительным), получим:

$$\varphi(y)=12y^3-19y^2-18y-7=0.$$

Так как $\varphi(2)<0$, а $\varphi(3)>0$, то $2<y_1<3$, и следует сделать подстановку $y=2+\frac{1}{z}$:

	12	-19	-18	-7
2	12	5	-8	-23
2	12	29	50	
2	12	53		

Отсюда

$$\psi(z)=23z^3-50z^2-53z-12=0.$$

Так как $\psi(3)=0$, то $z_1=3$, следовательно,

$$x_1=1+\frac{1}{2+\frac{1}{3}}=\frac{10}{7}.$$

Пример 2. Извлечь $\sqrt[3]{9}$ с точностью до 10^{-5}. Иными словами, надо найти единственный действительный корень уравнения $x^3-9=0$.

Очевидно, $2<x_1<3$, следовательно, надо сделать подстановку $x=2+\frac{1}{y}$:

	1	0	0	−9
2	1	2	4	−1
2	1	4	12	
2	1	6		

Следовательно, уравнение для y будет:

$$\varphi(y)=y^3-12y^2-6y-1=0.$$

Левую часть удобнее представить в виде

$$\varphi(y)=y^2\left(y-12-\frac{6}{y}-\frac{1}{y^2}\right),$$

так как при таком представлении сразу видно, что $\varphi(12)<0$, а $\varphi(13)>0$, т. е. $12<y_1<13$; делаем подстановку $y=12+\frac{1}{z}$:

	1	−12	−6	−1
12	1	0	−6	−73
12	1	12	138	
12	1	24		

Уравнение для z будет иметь вид

$$\psi(z)=73z^3-138z^2-24z-1=z^2\left(73z-138-\frac{24}{z}-\frac{1}{z^2}\right)=0.$$

Сразу видно, что $\psi(2)<0$, $\psi(3)>0$, т. е. $2<z_1<3$; делаем подстановку $z=2+\frac{1}{u}$:

	73	−138	−24	−1
2	73	8	−8	−17
2	73	154	300	
2	73	300		

Очередное уравнение

$$F(u)=u^2\left(17u-300-\frac{300}{u}-\frac{73}{u^2}\right)=0.$$

Легко видеть, что $F(18)<0$, а $F(19)>0$, т. е. $18<u_1<19$.

Если параллельно с приведёнными выше вычислениями вычислять соответствующие подходящие дроби для

$$x_1 = 2 + \cfrac{1}{12 + \cfrac{1}{2 + \cfrac{1}{18 + \ddots}}}$$

k	0	1	2	3	4
P_k	2	25	52	961	...
Q_k	1	12	25	462	...
q_k	2	12	2	18	...

и подсчитывать каждый раз, чему равно выражение $Q_k(Q_k + Q_{k-1})$, то легко установить, что $\frac{P_3}{Q_3} = \frac{961}{462} = 2{,}080086\ldots$ представляет искомый корень с точностью, даже превышающей требуемую, так как

$$Q_3(Q_3 + Q_2) = 462 \cdot 487 > 10^5 = \frac{1}{\alpha}.$$

В силу свойства II $\frac{P_3}{Q_3}$ является для искомого корня приближённым значением с избытком, а так как $\frac{1}{462 \cdot 487} < 0{,}000005$, то можно написать, что

$$2{,}080081 < \sqrt[3]{9} < 2{,}080087$$

или

$$\sqrt[3]{9} = 2{,}080083\ (\pm 3 \cdot 10^{-6})$$

(из таблиц имеем $\sqrt[3]{9} \approx 2{,}0800838$).

§ 6. Способ Лобачевского

Нашему великому соотечественнику Н. И. Лобачевскому принадлежит способ решения алгебраических уравнений, являющийся одним из самых удобных, особенно в тех случаях, когда надо определить и действительные, и мнимые корни уравнения [1]).

[1]) Способ Н. И. Лобачевского был открыт им не позднее 1832 г. Он изложен в его сочинении «Алгебра или вычисление конечных», вышедшем в свет в 1834 г. (Новое издание этого сочинения см.: Н. И. Лобачевский, Полное собрание сочинений, т. IV, Гостехиздат, М.—Л., 1948). Независимо от Н. И. Лобачевского аналогичный способ был предложен бельгийским математиком Данделеном (1826) и швейцарским математиком Греффе (1837). Работа последнего получила наибольшее распространение, почему в учебной литературе способ этот долгое время назывался по имени Греффе.

Одно из существенных удобств этого способа заключается в том, что нет необходимости предварительно определять границы действительных корней, общее число их, отделять корни друг от друга; всё это, как правило, автоматически определяется в процессе весьма несложных вычислений, сводящихся в основном к сложению, вычитанию и умножению чисел и дающих сразу все корни уравнения. Перейдём к изложению способа.

Пусть дано уравнение с действительными коэффициентами

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0, \tag{1}$$

имеющее своими корнями числа

$$x_1,\ x_2,\ x_3,\ \ldots,\ x_n.$$

Тогда, во-первых, составляется уравнение

$$f_1(x)=0 \tag{2}$$

(способ его составления будет описан ниже), корнями которого будут числа

$$-x_1^2,\ -x_2^2,\ \ldots,\ -x_n^2,$$

т. е. квадраты корней уравнения (1), взятые со знаком минус.

Затем с помощью того же преобразования переходят от уравнения (2) к уравнению

$$f_2(x)=0, \tag{3}$$

корнями которого будут числа:

$$-(-x_1^2)^2=-x_1^4,\ -x_2^4,\ \ldots,\ -x_n^4.$$

После k-кратного повторения этого преобразования приходят к уравнению

$$f_k(x)=A_0x^n+A_1x^{n-1}+\ldots+A_n=0, \tag{4}$$

корнями которого будут числа

$$-x_1^{2^k},\ -x_2^{2^k},\ \ldots,\ -x_n^{2^k}.$$

Смысл этих преобразований особенно отчётливо усматривается, когда все корни уравнения (1) действительны и различны по

абсолютной величине; действительно, пусть

$$|x_1|>|x_2|>\ldots>|x_{n-1}|>|x_n|. \tag{5}$$

Введём для краткости обозначение

$$2^k=s \tag{6}$$

и напишем соотношения Вьета для уравнения (4):

$$\left.\begin{aligned}
\frac{A_1}{A_0}&=x_1^s+x_2^s+x_3^s+\ldots+x_n^s=\\
&=x_1^s\left[1+\left(\frac{x_2}{x_1}\right)^s+\ldots+\left(\frac{x_n}{x_1}\right)^s\right]=x_1^s(1+\alpha_1),\\
\frac{A_2}{A_0}&=x_1^sx_2^s+x_1^sx_3^s+x_2^sx_3^s+x_2^sx_4^s+\ldots=\\
&=x_1^sx_2^s\left[1+\left(\frac{x_3}{x_2}\right)^s+\left(\frac{x_3}{x_1}\right)^s+\ldots\right]=x_1^sx_2^s(1+\alpha_2),\\
\frac{A_3}{A_0}&=x_1^sx_2^sx_3^s+x_1^sx_2^sx_4^s+x_2^sx_3^sx_4^s+\ldots=\\
&=x_1^sx_2^sx_3^s\left[1+\left(\frac{x_4}{x_3}\right)^s+\ldots\right]=x_1^sx_2^sx_3^s(1+\alpha_3),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\frac{A_{n-1}}{A_0}&=x_1^sx_2^s\ldots x_{n-1}^s+x_1^s\ldots x_{n-2}^sx_n^s+\ldots=\\
&=x_1^sx_2^s\ldots x_{n-1}^s\left[1+\left(\frac{x_n}{x_{n-1}}\right)^s+\ldots\right]=\\
&=x_1^sx_2^s\ldots x_{n-1}^s(1+\alpha_{n-1}),\\
\frac{A_n}{A_0}&=x_1^sx_2^s\ldots x_{n-1}^sx_n^s.
\end{aligned}\right\} \tag{7}$$

Деля почленно каждое из соотношений (7) (начиная со второго) на предыдущее, получим:

$$\left.\begin{aligned}
x_1^s&=\frac{A_1}{A_0(1+\alpha_1)},\\
x_2^s&=\frac{A_2(1+\alpha_1)}{A_1(1+\alpha_2)},\\
&\ldots\ldots\ldots\\
x_n^s&=\frac{A_n(1+\alpha_{n-1})}{A_{n-1}}.
\end{aligned}\right\} \tag{8}$$

При условии (5) числа $\frac{x_2}{x_1}$, $\frac{x_3}{x_1}$, $\frac{x_3}{x_2}$ и т. п., вообще числа вида $\frac{x_l}{x_m}$ $(l>m)$ будут по абсолютной величине меньше единицы; поэтому при достаточно больших k, а следовательно и s, положительные числа α_1, α_2, ... , α_{n-1} будут как угодно малыми.

Поэтому соотношения (8) можно заменить приближёнными соотношениями

$$\left.\begin{aligned} x_1^s &\approx \frac{A_1}{A_0}, \\ x_2^s &\approx \frac{A_2}{A_1}, \\ &\cdots\cdots\cdots \\ x_n^s &\approx \frac{A_n}{A_{n-1}} \end{aligned}\right\} \qquad (9)$$

с как угодно малыми (при больших значениях s) относительными погрешностями.

Таким образом, корни уравнения (4) благодаря резкому отличию их друг от друга могут быть определены с большой точностью по формулам (9); после же извлечения с помощью таблиц логарифмов корней степени s из чисел (9) получим и корни уравнения (1) (надо лишь для каждого m определить, будет ли уравнению (1) удовлетворять $+\sqrt[s]{\frac{A_m}{A_{m-1}}}$ или $-\sqrt[s]{\frac{A_m}{A_{m-1}}}$).

Перейдём теперь к рассмотрению вопроса о переходе от уравнения (1) к уравнению (2), затем (3) и т. д.

Для вывода правила составления функции $f_1(x)$, корни которой были бы равны $-x_1^2, -x_2^2, \ldots, -x_n^2$, удобнее всего пойти следующим, несколько искусственным путём.

Для многочлена $f(x)$ имеем:

$$a_0(x-x_1)(x-x_2)\ldots(x-x_n) \equiv$$
$$\equiv a_0x^n + a_1x^{n-1} + a_2x^{n-2} + a_3x^{n-3} + \ldots + a_{n-2}x^2 + a_{n-1}x + a_n. \qquad (10)$$

Если в этом тождестве заменим x на $-x$ и обе части тождества умножим на $(-1)^n$ (чтобы в правой части старший коэффициент сохранил свой знак), то получим:

$$a_0(x+x_1)(x+x_2)\ldots(x+x_n) \equiv$$
$$\equiv a_0x^n - a_1x^{n-1} + a_2x^{n-2} - a_3x^{n-3} + \ldots + (-1)^{n-2}a_{n-2}x^2 +$$
$$+ (-1)^{n-1}a_{n-1}x + (-1)^n a_n. \qquad (11)$$

Перемножая почленно тождества (10) и (11), получим:

$$a_0^2(x^2-x_1^2)(x^2-x_2^2)\ldots(x^2-x_n^2) \equiv a_0^2x^{2n} - (a_1^2 - 2a_0a_2)x^{2n-2} +$$
$$+ (a_2^2 - 2a_1a_3 + 2a_0a_4)x^{2n-4} - \ldots + (-1)^k(a_k^2 - 2a_{k-1}a_{k+1} +$$
$$+ 2a_{k-2}a_{k+2} - \ldots)x^{2n-2k} + \ldots + (-1)^{n-1}(a_{n-1}^2 - 2a_{n-2}a_n)x^2 +$$
$$+ (-1)^n a_n^2. \qquad (12)$$

Заменяя, наконец, в тождестве (12) x^2 на $-x$ и умножая обе части его на $(-1)^n$, получим:

$$a_0^2(x+x_1^2)(x+x_2^2)\ldots(x+x_n^2)\equiv a_0^2x^n+(a_1^2-2a_0a_2)x^{n-1}+\ldots+$$
$$+(a_k^2-2a_{k-1}a_{k+1}+\ldots)x^{n-k}+\ldots+(a_{n-1}^2-2a_{n-2}a_n)x+a_n^2. \quad (13)$$

Стоящий здесь справа многочлен и есть искомая функция $f_1(x)$, так как из тождества (13) видно, что он обращается в нуль при $x=-x_1^2, -x_2^2, \ldots, -x_n^2$.

Итак, если $x_1, x_2, \ldots, x_n$ — корни уравнения (1), то любой коэффициент уравнения, корни которого равны числам $-x_1^2$, $-x_2^2, \ldots, -x_n^2$, равен квадрату соответствующего коэффициента данного уравнения минус удвоенное произведение соседних с ним коэффициентов, плюс удвоенное произведение коэффициентов, индексы которых отличаются от данного на две единицы, минус удвоенное произведение коэффициентов, индексы которых отличаются от данного на три единицы, и т. д. до тех пор, пока не будет использован какой-нибудь из крайних коэффициентов.

Так как через два-три шага коэффициенты делаются очень большими, то дальнейшие вычисления приходится вести либо с помощью таблиц логарифмов, причём выгодно привлекать к вычислениям также и гауссовы таблицы логарифмов сумм и разностей [1]), либо на арифмометре, округляя каждый раз результаты вычислений так, чтобы у крупнейших слагаемых сохранялось определённое, зависящее от требуемой точности число значащих цифр; в последнем случае результаты обычно записываются в виде произведения десятичных дробей с целой частью, не превышающей десяти, на ту или другую степень десяти.

Запись вычислений удобно располагать так, как указано в проделанном ниже примере: помимо окончательных результатов, выписываются удвоенные произведения соответствующих «соседних» коэффициентов, что облегчает в случае необходимости проверку вычислений.

Удвоенные произведения будем выписывать всегда с учётом соответствующих знаков, т. е., например, будем писать числа a_2^2, $-2a_1a_3$, $+2a_0a_4$ и т. п., которые после этого остаётся лишь сложить, для чего удобно иметь под руками счёты; если какое-нибудь удвоенное произведение будет исчезающе мало (в пределах точности вычислений) по сравнению с квадратом соответствующего коэффициента, то вместо него будем ставить звёздочку (*).

Пример. Дано уравнение

$$f(x)=2x^3-15x^2+16x+12=0.$$

[1]) См., например, Пржевальский, Пятизначные таблицы логарифмов, стр. 155—171.

Составить функции $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_4(x)$, сохраняя при вычислениях четыре значащие цифры.

k	s	Коэффициенты уравнений			
Данное уравнение		2	—15	16	12
			$(-15)^2 = 225$ $-2 \cdot 2 \cdot 16 = -64$	$16^2 = 256$ $-2 \cdot (-15) \cdot 12 = 360$	
1	2	4	161	616	144
			$161^2 = 2{,}592 \cdot 10^4$ $-2 \cdot 4 \cdot 616 = -0{,}493 \cdot 10^4$	$616^2 = 3{,}795 \cdot 10^5$ $-2 \cdot 161 \cdot 144 =$ $= -0{,}464 \cdot 10^5$	
2	4	16	$2{,}099 \cdot 10^4$	$3{,}331 \cdot 10^5$	$2{,}074 \cdot 10^4$
			$4{,}406 \cdot 10^8$ $-0{,}107 \cdot 10^8$	$1{,}110 \cdot 10^{11}$ $-0{,}008 \cdot 10^{11}$	
3	8	256	$4{,}299 \cdot 10^8$	$1{,}102 \cdot 10^{11}$	$4{,}301 \cdot 10^8$
			$1{,}848 \cdot 10^{17}$ $-0{,}001 \cdot 10^{17}$	$1{,}214 \cdot 10^{22}$ *	
4	16	$6{,}554 \cdot 10^4$	$1{,}847 \cdot 10^{17}$	$1{,}214 \cdot 10^{22}$	$1{,}850 \cdot 10^{17}$

Из таблицы следует, что

$$f_4(x) = 6{,}554 \cdot 10^4 x^3 + 1{,}847 \cdot 10^{17} x^2 + 1{,}214 \cdot 10^{22} x + 1{,}850 \cdot 10^{17}.$$

Конечно, чем резче отличаются друг от друга корни уравнения (1), т. е. чем меньше будут величины $\left|\frac{x_m}{x_{m-1}}\right|$, тем быстрее будут убывать с возрастанием k величины $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ в соотношениях (7) и тем меньше надо сделать преобразований для достижения заданной точности. Если бы мы знали заранее верхнюю границу для чисел $\left|\frac{x_m}{x_{m-1}}\right|$, то смогли бы сразу для любого k оценить величины $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$.

Важно, однако, найти признак, по которому можно было бы судить, даже ничего не зная о величине чисел $\left|\frac{x_m}{x_{m-1}}\right|$, с какой точностью получатся корни данного уравнения после определённого числа преобразований и, наоборот, в какой момент можно, прекратив преобразования, применять формулы (9), если нам заданы границы погрешностей искомых корней. Для установления этого признака предположим, что мы ещё раз преобразовали уравнение (4), перейдя к уравнению

$$f_{k+1}(x) = B_0 x^n + B_1 x^{n-1} + \ldots + B_n = 0, \tag{14}$$

корни которого равны $-x_1^{2s}, -x_2^{2s}, \ldots, -x_n^{2s}$ $(2^{k+1} = 2 \cdot 2^k = 2s)$. Аналогично формулам (7) будем иметь:

$$\frac{B_m}{B_0} = x_1^{2s} x_2^{2s}, \ \ldots, \ x_m^{2s}(1 + \beta_m) \tag{15}$$

(здесь m может быть любым числом от единицы до n), причём положительные величины $\beta_1, \beta_2, \ldots, \beta_{n-1}$ будут меньше соответствующих величин α_i, а $\beta_n = 0$.

Так как

$$\frac{A_m^2}{A_0^2} = [x_1^s x_2^s \ldots x_m^s (1 + \alpha_m)]^2, \tag{16}$$

а $B_0 = A_0^2$, то, деля почленно соотношение (15) на (16), получим:

$$B_m = A_m^2 \frac{1 + \beta_m}{(1 + \alpha_m)^2}. \tag{17}$$

Соотношение (17) показывает, что в том случае, когда величины $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ в формулах (7) достаточно малы по сравнению с единицей, так что в пределах точности вычислений соотношения (8) могут быть заменены приближёнными соотношениями (9), с той же примерно точностью будем иметь:

$$B_m \approx A_m^2. \tag{17'}$$

Таким образом, преобразование уравнений мы должны вести до тех пор, пока все удвоенные произведения «соседних коэффициентов» не сделаются исчезающе малыми (в пределах точности вычислений) по сравнению с квадратами соответствующих коэффициентов.

Отметим без доказательства, что справедливо и обратное утверждение: когда для всех значений m будут иметь место (в пределах точности вычислений) соотношения $B_m \approx A_m^2$, то с той же точностью будут справедливы равенства (9), т. е. можно, прекратив дальнейшие преобразования, приступить к вычислению корней по формулам (9).

Правда, с течением времени накапливается вычислительная погрешность, особенно при большом числе преобразований. Поэтому нельзя быть уверенным, что окончательные значения корней будут иметь столько же верных цифр, сколько сохранялось при вычислениях; однако можно рассчитывать, вообще говоря, на то, что число верных цифр в окончательных ответах будет лишь на единицу меньше, чем число цифр, сохранявшихся при вычислениях.

Отметим, что когда все корни уравнения (1) действительны, то уравнения (2), (3) и т. п. имеют исключительно отрицательные корни, откуда следует, что все коэффициенты будут обязательно положительными (в чём легко убедиться, применив теорему Декарта — см. стр. 331).

Таким образом, если какое-нибудь из уравнений (2), (3) и т. п. будет иметь хотя бы один отрицательный коэффициент, то можно утверждать, что данное уравнение (1) имеет мнимые корни.

Вычисление мнимых корней. Рассмотрим сначала простейший случай, когда уравнение (1)

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0$$

имеет лишь одну пару мнимых корней, например:

$$x_{2,\,3}=u\pm vi=r(\cos\varphi\pm i\sin\varphi), \tag{18}$$

причём

$$|x_1|>r>|x_4|>\ldots>|x_{n-1}|>|x_n|. \tag{19}$$

Напомним, что сумма и произведение одинаковых степеней двух сопряжённых комплексных чисел будут числами действительными:

$$x_2^s+x_3^s=2r^s\cos s\varphi,\quad x_2^sx_3^s=r^{2s}. \tag{20}$$

При условиях (19) в любом из равенств (7), кроме второго, наиболее крупным модулем будет попрежнему обладать первое слагаемое, а потому будем иметь:

$$\frac{A_1}{A_0}=x_1^s(1+\alpha_1),\ \frac{A_3}{A_0}=x_1^sr^{2s}(1+\alpha_3),\ \frac{A_4}{A_0}=x_1^sr^{2s}x_4^s(1+\alpha_4),\ldots \tag{21}$$

В силу равенств (20) второе из соотношений (7) нам даст:

$$\begin{aligned}\frac{A_2}{A_0}&=(x_1^sx_2^s+x_1^sx_3^s)+x_2^sx_3^s+x_1^sx_4^s+x_2^sx_4^s+\ldots=\\&=2x_1^sr^s\cos s\varphi+r^{2s}+x_1^sx_4^s+\ldots=\\&=2x_1^sr^s\left(\cos s\varphi+\frac{r}{2x_1^s}+\frac{x_4^s}{2r^s}+\ldots\right).\end{aligned} \tag{22}$$

Модули всех стоящих в скобках слагаемых, кроме $\cos s\varphi$, будут при возрастании s стремиться к нулю, при этом сумма их [так же

как и величины $\alpha_1, \alpha_3, \ldots$ в (21)] будет числом действительным, так как комплексные слагаемые входят в неё сопряжёнными парами. Что же касается $\cos s\varphi$, то его изменение будет зависеть от угла φ. Например, при $\varphi = 120°$

$$\cos(2^k\varphi) = \cos s\varphi = -\frac{1}{2};$$

если $\varphi = \frac{\pi}{2^m}$, то при $s = 2^k > 2^m$ будем иметь: $\cos s\varphi = 1$; вообще же $\cos s\varphi$ при своём изменении будет принимать и положительные и отрицательные значения как очень малые по абсолютной величине, так и близкие к единице.

Но даже в случае чисто мнимых корней $\left(\varphi = \frac{\pi}{2}\right)$, когда для второго и следующих преобразованных уравнений $\cos s\varphi$ будет равен 1, коэффициент при x^{n-2} не будет меняться даже при больших значениях s так же правильно, как при действительных корнях. Действительно, из равенств

$$\frac{A_2}{A_0} = 2x_1^s r^s (1 + \alpha_2) \quad \text{и} \quad \frac{B_2}{B_0} = 2x_1^{2s} r^{2s} (1 + \beta_2)$$

будет следовать:

$$\frac{A_2^2}{A_0^2} \approx 4x_1^{2s} r^{2s} \approx 2\frac{B_2}{B_0},$$

а так как $B_0 = A_0^2$, то

$$B_2 \approx \frac{1}{2} A_2^2. \tag{23}$$

Однако неправильное изменение коэффициента при x^{n-2} не мешает из соотношений (21) получить:

$$x_1^s \approx \frac{A_1}{A_0}, \quad r^{2s} \approx \frac{A_3}{A_1}, \quad x_4^s \approx \frac{A_4}{A_3}, \quad \ldots, \quad x_n^s \approx \frac{A_n}{A_{n-1}}, \tag{24}$$

что позволяет после извлечения корней (и выбора подходящих знаков) определить

$$x_1, x_4, x_5, \ldots, x_n$$

и

$$r^2 = +\sqrt[s]{\frac{A_3}{A_1}}.$$

Но

$$x_1 + (x_2 + x_3) + \ldots + x_n = x_1 + 2u + x_4 + \ldots + x_n = -\frac{a_1}{a_0}, \tag{25}$$

откуда легко найти действительную часть искомых мнимых корней

$$u = -\frac{1}{2}\left(\frac{a_1}{a_0} + x_1 + x_4 + \ldots + x_n\right), \tag{26}$$

а затем и мнимую часть

$$v = \sqrt{r^2 - u^2}. \tag{27}$$

Поясним примером применение формул (24), (26), (27).

Пример. Решить уравнение

$$12x^3 - 17x^2 + 55x + 15 = 0.$$

k	s	Коэффициенты уравнений			
		12	-17	55	15
			289 -1320	3025 510	
1	2	144	$-1{,}03\cdot10^3$	$3{,}54\cdot10^3$	225
			$1{,}06\cdot10^6$ $-1{,}02\cdot10^6$	$1{,}25\cdot10^7$ $+0{,}05\cdot10^7$	
2	4	$2{,}07\cdot10^4$	$0{,}04\cdot10^6$	$1{,}30\cdot10^7$	$5{,}06\cdot10^4$

Легко видеть, что в следующем преобразованном уравнении все коэффициенты, кроме второго, будут квадратами полученных чисел (в пределах трёх значащих цифр, сохраняемых при вычислениях). Так как в преобразованных уравнениях встречались отрицательные коэффициенты, то, несомненно, данное уравнение имеет мнимые корни, причём

$$x_{1,2} = r(\cos\varphi \pm i\sin\varphi) = u \pm vi \quad \text{и} \quad r > |x_3|,$$

потому что неправильно меняются коэффициенты при x^{n-1}, а не при x^{n-2}, как было в разобранном нами случае.

Итак,

$$r^2 \approx \sqrt[4]{\frac{1{,}30\cdot10^7}{2{,}07\cdot10^4}} \approx 5{,}006, \quad x_3 \approx \sqrt[4]{\frac{5{,}06\cdot10^4}{1{,}30\cdot10^7}} \approx -0{,}2498$$

(по теореме Декарта данное уравнение или совсем не имеет положительных корней, или имеет их два; но при двух мнимых корнях у уравнения имеется один лишь действительный корень; следовательно, он должен быть отрицательным).

По формулам (26) и (27) имеем:

$$u \approx -\frac{1}{2}\left(-\frac{17}{12} - 0{,}2498\right) \approx 0{,}8333, \quad v \approx \sqrt{5{,}006 - 0{,}8333^2} \approx 2{,}077.$$

Следовательно, $x_{1,2} \approx 0{,}8333 \pm 2{,}077\,i$ (точные значения корней

$$x_3 = -0{,}25, \quad x_{1,2} = \frac{5}{6} \pm i\cdot\frac{\sqrt{155}}{6} = 0{,}83333\ldots \pm i\cdot 2{,}0742\ldots).$$

Легко убедиться с помощью соотношений (7), что при двух парах мнимых корней, например при

$$x_{1,2} = r_1(\cos\varphi_1 \pm i\sin\varphi_1) = u_1 \pm v_1 i$$

и

$$x_{5,6} = r_2(\cos\varphi_2 \pm i\sin\varphi_2) = u_2 \pm v_2 i,$$

если при этом $r_1 > |x_3| > |x_4| > r_2 > |x_5| > \ldots > |x_n|$, неправильно себя будут вести коэффициенты при x^{n-1} и при x^{n-3}; остальные же коэффициенты, начиная с некоторого момента, будут (в пределах точности вычислений) просто возводиться в квадрат при очередных преобразованиях; вместо (24) будем иметь:

$$\left.\begin{aligned} r_1^2 \approx \sqrt[s]{\frac{A_2}{A_0}},\quad x_3 \approx \sqrt[s]{\frac{A_3}{A_2}},\quad x_4 \approx \sqrt[s]{\frac{A_4}{A_3}},\\ r_2^2 \approx \sqrt[s]{\frac{A_5}{A_4}},\quad x_7 \approx \sqrt[s]{\frac{A_7}{A_6}},\quad \ldots \end{aligned}\right\} \tag{28}$$

Для определения u_1 и u_2 следует воспользоваться соотношениями

$$\left.\begin{aligned} x_1 + x_2 + \ldots + x_n = 2u_1 + x_3 + x_4 + 2u_2 + x_7 + \ldots = -\frac{a_1}{a_0},\\ \frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} = \frac{1}{u_1 + v_1 i} + \frac{1}{u_1 - v_1 i} + \frac{1}{x_3} + \ldots = -\frac{a_{n-1}}{a_n} \end{aligned}\right\} \tag{29}$$

(последнее соотношение следует из того, что корни уравнения $a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = 0$ равны $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \ldots, \frac{1}{x_n}$). Так как $\frac{1}{u_1 + v_1 i} + \frac{1}{u_1 - v_1 i} = \frac{2u_1}{u_1^2 + v_1^2} = \frac{2u_1}{r_1^2}$ и аналогично для второй пары мнимых корней, то соотношения (29) дают:

$$\left.\begin{aligned} 2u_1 + 2u_2 = -\left(\frac{a_1}{a_0} + x_3 + x_4 + x_7 + \ldots + x_n\right),\\ \frac{2u_1}{r_1^2} + \frac{2u_2}{r_2^2} = -\left(\frac{a_{n-1}}{a_n} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{1}{x_7} + \ldots + \frac{1}{x_n}\right), \end{aligned}\right\} \tag{29'}$$

откуда легко найти u_1, u_2, а затем и v_1, v_2 по формулам

$$v_1 = \sqrt{r_1^2 - u_1^2},\quad v_2 = \sqrt{r_2^2 - u_2^2}. \tag{30}$$

Случай действительных корней, почти одинаковых по абсолютной величине. В том случае, когда все корни уравнения (1) действительны, но вместо (5) имеем, например,

$$|x_1| > |x_2| \approx |x_3| > |x_4| > \ldots > |x_n|, \tag{31}$$

второе из соотношений (7) даёт:

$$\frac{A_2}{A_0} = x_1^s(x_2^s + x_3^s) + x_2^s x_3^s + x_1^s x_4^s + \ldots \approx$$
$$\approx 2x_1^s x_2^s + x_2^{2s} + x_1^s x_4^s + \ldots = 2x_1^s x_2^s[1 + z_2'].$$

Если s настолько велико, что α_2' по сравнению с единицей будет весьма малой величиной, то при следующем преобразовании получим уравнение вида (14), где

$$B_2 \approx \frac{1}{2} A_2^2, \tag{23'}$$

в чём легко убедиться, рассуждая так же, как и при выводе соотношения (23).

Игнорируя неправильно изменяющийся коэффициент, из равенств

$$\frac{A_1}{A_0} \approx x_1^s, \quad \frac{A_3}{A_0} \approx x_1^s x_2^s x_3^s \approx x_1^s x_2^{2s}, \quad \frac{A_4}{A_0} \approx x_1^s x_2^s x_3^s x_4^s, \ldots,$$

получим:

$$x_1 \approx \sqrt[s]{\frac{A_1}{A_0}}, \quad x_{2,3} \approx \sqrt[2s]{\frac{A_3}{A_1}}, \quad x_4 \approx \sqrt[s]{\frac{A_4}{A_3}}, \ldots$$

Как всегда, после извлечения корней надо проверить, какой из знаков будет подходящим для данного уравнения. При этом может оказаться, что x_2 и x_3 будут разных знаков (т. е. они почти совпадают только по абсолютной величине); но возможно, что x_2 и x_3 будут иметь одинаковые знаки (случай кратного или «почти кратного» корня).

Как мы видели выше, в случае мнимых $x_{2,3}$ при $\varphi = \frac{\pi}{2m}$ поведение коэффициента при x^{n-2} может сделаться похожим на поведение этого коэффициента при x_2, x_3 действительных и $|x_2| \approx |x_3|$.

Следует заметить, что для распознавания мнимых корней служит, вообще говоря, наличие отрицательных коэффициентов в каком-нибудь из преобразованных уравнений.

Правда, может случиться (например, при очень малом φ), что отрицательные знаки не успеют появиться, а мы обнаружим изменение какого-нибудь коэффициента, характерное для двух действительных корней, имеющих почти равные абсолютные значения.

Пример. Решить уравнение

$$f(x) = x^3 + 101x^2 + 2601x + 2501 = (x^2 + 100x + 2501)(x + 1) = 0,$$

корни которого

$$x_{1,2} = -50 \pm i \approx 50{,}01\,(\cos 1°10' \pm i \sin 1°10'), \qquad x_3 = -1.$$

Так как дальнейшее изменение коэффициента при x и свободного члена будет правильным, а изменение коэффициента при x^2 подчиняется соотношению (23), и, кроме того, ни одно из преобразованных уравнений не содержит отрицательных коэффициентов, можно было бы предположить, что мы имеем случай двух близких по абсолютной величине корней:

$$x_{1,2} \approx \sqrt[8]{\frac{3{,}92 \cdot 10^{13}}{1}} \approx \pm 50, \quad x_3 \approx \sqrt[4]{\frac{3{,}92 \cdot 10^{13}}{3{,}92 \cdot 10^{13}}} = \pm 1.$$

k	s	Коэффициенты уравнений			
		1	101	2601	2501
			$1{,}02 \cdot 10^4$ $-0{,}52 \cdot 10^4$	$6{,}77 \cdot 10^6$ $-0{,}50 \cdot 10^6$	
1	2	1	$0{,}50 \cdot 10^4$	$6{,}27 \cdot 10^6$	$6{,}26 \cdot 10^3$
			$2{,}50 \cdot 10^7$ $-1{,}25 \cdot 10^7$	$3{,}93 \cdot 10^{13}$ $-0{,}01 \cdot 10^{13}$	
2	4	1	$1{,}25 \cdot 10^7$	$3{,}92 \cdot 10^{13}$	$3{,}92 \cdot 10^{13}$

Конечно, — 1 удовлетворяет нашему уравнению, но ни 50, ни — 50 и никакое близкое к ним действительное число корнем данного уравнения не будут.

Разделив $f(x)$ на $x+1$, можно из соответствующего квадратного уравнения искать мнимые корни.

Можно было бы поступить и иначе: вести вычисления не с тремя значащими цифрами, а, например, с семью цифрами (при наличии арифмометра это легко выполнимо). Тогда через несколько преобразований появились бы отрицательные коэффициенты, свидетельствующие о мнимых корнях.

Конечно, в подобного рода случаях может быть использована и теорема Штурма.

В случаях, когда уравнение имеет три пары мнимых корней, особенно, когда модули у двух или трёх пар корней почти одинаковы или когда модуль какой-нибудь пары совпадает с абсолютным значением действительного корня и т. п., вычисления по способу Лобачевского усложняются. Отсылая интересующихся подробностями к специальной литературе, особо укажем на «Лекции о приближённых вычислениях» акад. А. Н. Крылова, где подробно рассмотрены различные возможности и указаны, в частности, разные приёмы вычисления аргументов φ_1, φ_2, φ_3, ... мнимых корней, приёмы уточнения полученных по способу Лобачевского корней и т. п.

Задачи к главе I

1. Отделить по способу Штурма корни уравнений:

а) $x^3 + x - 24 = 0$,

б) $x^3 - 60x + 4 = 0$,

в) $x^4 - 6x^2 + 4 = 0$,

найдя для каждого корня интервал, длина которого не превышала бы единицы.

2. Найти границы действительных корней уравнений:

а) $x^5 + 2x^4 - 10x^3 + 20x - 40 = 0,$

б) $x^6 + x^4 - x^3 - 4x^2 + 60 = 0,$

в) $4x^5 - x^2 + 20x - 2 = 0.$

3. Вычислить с точностью до 0,01 по способу Горнера и по способу Лагранжа положительные корни уравнений:

а) $x^3 - 28 = 0,$

б) $x^3 - 15x + 4 = 0,$

в) $x^4 + x - 19 = 0.$

4. Отделить по способу Штурма и вычислить по способу Лагранжа и по способу Горнера с точностью до 0,01 корни уравнений:

а) $15x^3 - 65x^2 + 93x - 44 = 0,$

б) $3x^3 + 4x^2 - 9x - 12 = 0.$

5. Доказать, что нижняя граница положительных корней уравнения равна $\frac{1}{K_2}$, а верхняя граница отрицательных корней равна $-\frac{1}{K_3}$ (см. конец § 2).

6. Найти по способу Горнера с точностью до 0,0001 наименьшее положительное значение $\frac{x}{R}$ из уравнения

$$x^3 - 3R^2x + R^3 = 0$$

и истолковать геометрический смысл результата [см. уравнение (5) во введении].

7. Используя идею способа Лагранжа, найти первые три неполных частных разложения в непрерывную дробь корня уравнения $10^x = 2$.

8. Доказать, что, когда все корни уравнения $f(x) = 0$ действительны и $|x_1| \approx |x_2| \approx |x_3| > |x_4| > \ldots > |x_n|$, в способе Лобачевского неправильно будут меняться коэффициенты при x^{n-1} и при x^{n-2}; при этом

$$|x_1| \approx |x_2| \approx |x_3| \approx \sqrt[3s]{\frac{A_3}{A_0}}, \quad |x_4| \approx \sqrt[s]{\frac{A_4}{A_3}} \text{ и т. д.}$$

9. Решить по способу Лобачевского уравнения:

а) $3x^3 - 32x^2 + 170x - 100 = 0,$

б) $x^4 - 9x^3 + 41x^2 + 40x + 50 = 0,$

в) $3x^3 - 41x^2 + 113x - 11 = 0,$

г) $x^4 - 13x^3 + 27x^2 - 28x + 6 = 0.$

ГЛАВА II

РЕШЕНИЕ ТРАНСЦЕНДЕНТНЫХ УРАВНЕНИЙ

§ 7. Способ линейного интерполирования и способ Ньютона

Переходя к решению трансцендентных уравнений, заметим, прежде всего, что любой из описанных ниже способов будет пригоден и для решения уравнений алгебраических.

Как было указано в предисловии, табличный способ решения уравнений позволяет уменьшать как угодно интервал, внутри которого лежит действительный корень данного уравнения (трансцендентного или алгебраического):

$$f(x)=0. \tag{1}$$

По своей идее табличный способ чрезвычайно прост, но каждое существенное уменьшение интервала, заключающего искомый корень, связано с большой вычислительной работой, так как требует вычисления большого числа значений функции $f(x)$.

Значительно быстрее можно достигнуть цели при комбинированном использовании двух способов: способа линейного интерполирования, называемого также *правилом ложного положения* (regula falsi) [1]), и способа Ньютона. Перейдём к их описанию.

Пусть в интервале (a, b) лежит лишь один действительный корень x_1 уравнения (1), причём знаки $f(a)$ и $f(b)$ различны.

Пусть, кроме того,

$$f''(x) \text{ в интервале } (a, b) \text{ не обращается в нуль.} \tag{2}$$

Так как мы предполагаем $f''(x)$ непрерывной, то отсюда следует, что кривая, представляющая график функции $f(x)$, либо обращена своей вогнутостью вверх на протяжении всего интервала (при $f''(x)>0$), либо — вниз (при $f''(x)<0$).

Обе эти возможности изображены на рис. 1, где буквой a обозначен тот конец интервала, где $f(x)$ и $f''(x)$ имеют одинаковые знаки, а не обязательно левый конец, как обычно принято.

---

[1]) Или правилом двух ложных положений (regula duorum falsorum).

Переход к более узкому интервалу (a_1, b_1) осуществляется так. Взяв уравнение прямой MN, проходящей через две точки $M[a, f(a)]$ и $N[b, f(b)]$:

$$\frac{y-f(a)}{f(b)-f(a)}=\frac{x-a}{b-a}, \tag{3}$$

определяют абсциссу точки пересечения MN и осью Ox, или то

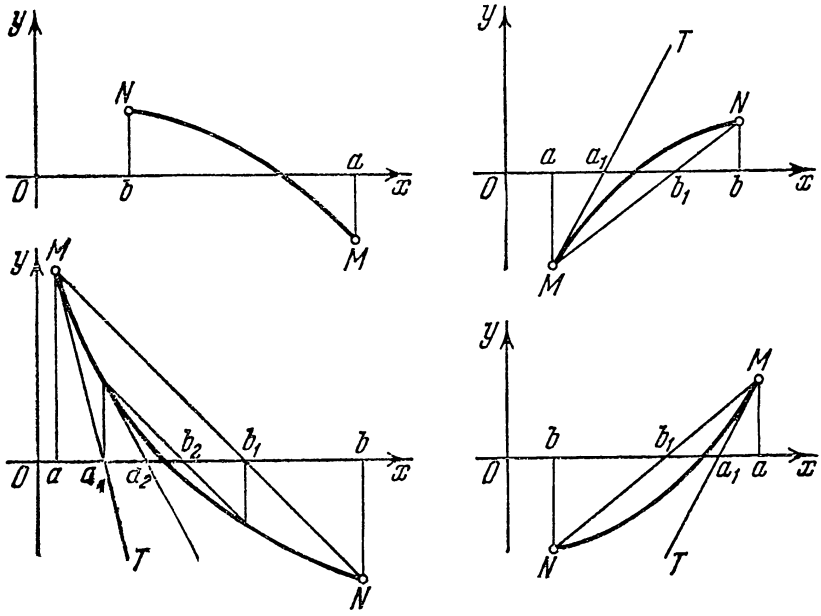

Рис. 1.

значение x, при котором y в уравнении (3) обращается в нуль, т. е. решают уравнение

$$\frac{0-f(a)}{f(b)-f(a)}=\frac{x-a}{b-a}.$$

Корень этого уравнения принимают за границу b_1 нового интервала; следовательно,

$$b_1=a-\frac{f(a)(b-a)}{f(b)-f(a)}. \tag{4}$$

Таким образом, функция $f(x)$ как бы заменяется линейной функцией (т. е. функцией первой степени):

$$y=f(a)+\frac{f(b)-f(a)}{b-a}(x-a), \tag{5}$$

определяемой из равенства (3).

Эта функция носит название *интерполяционного многочлена* и обладает тем свойством, что её значения совпадают со значениями функции $f(x)$ при $x=a$ и при $x=b$.

Отсюда и самый способ вычисления по формуле (4) границы b_1 нового, более узкого, интервала носит название *способа линейного интерполирования.*

Примечание. Напомним, что линейное интерполирование применяют также при пользовании различными таблицами, когда ищут значение функции для аргумента, лежащего между какими-нибудь двумя его значениями, приведёнными в таблице. Например, зная, что

$$\lg 1705 = 3{,}23172,$$
$$\lg 1706 = 3{,}23198$$

считают, что в интервале (1705, 1706) функция $\lg x$ может быть заменена интерполяционным многочленом первой степени вида (5), т. е.

$$\lg x \approx \lg 1705 + \frac{\lg 1706 - \lg 1705}{1706 - 1705}(x - 1705), \qquad (6)$$

и если надо найти, например, $\lg 1705{,}4$, то вычисляют значение многочлена (6):

$$\lg x \approx 3{,}23\,172 + 0{,}00026 \cdot 0{,}4 \approx 3{,}32182.$$

В тех случаях, когда такая замена грозит большой погрешностью, применяют так называемое параболическое интерполирование, т. е. заменяют данную функцию многочленом, например, второй степени, значения которого совпадали бы со значениями функции не в двух, а в трёх точках.

Однако для интересующей нас цели (получение b_1 — улучшенной границы для искомого корня) использование интерполяционных многочленов даже второй степени практически нецелесообразно.

Другая граница интервала (a, b) улучшается по способу Ньютона, согласно которому считают, что

$$f(x) \approx f(a) + f'(a)(x - a), \qquad (7)$$

т. е. пренебрегают в разложении $f(x)$ в ряд Тейлора членами, содержащими $(x-a)^2$, $(x-a)^3$ и т. д.

Приравнивая правую часть равенства (7) нулю:

$$f(a) + f'(a)(x - a) = 0,$$

принимают корень полученного уравнения за вторую границу a_1 нового интервала:

$$a_1 = a - \frac{f(a)}{f'(a)}. \qquad (8)$$

Так как уравнение касательной MT, проведённой в точке $M[a, f(a)]$ (см. рис. 1), имеет вид

$$y = f(a) + f'(a)(x - a),$$

то a_1, найденное по формуле (8), будет абсциссой точки пересечения касательной MT с осью Ox.

Если окажется, что новый интервал (a_1, b_1) всё ещё слишком велик, то, повторно применяя способ линейного интерполирования и способ Ньютона, вычисляют

$$b_2 = a_1 - \frac{f(a_1)(b_1 - a_1)}{f(b_1) - f(a_1)},$$

$$a_2 = a_1 - \frac{f(a_1)}{f'(a_1)}$$

и поступают так до тех пор, пока не получится интервал (a_k, b_k) достаточно малых размеров (см. рис. 1).

Иногда для приближения к исходному корню используют лишь способ Ньютона, вычисляя a_1, a_2, a_3 и т. д., что с геометрической точки зрения равносильно проведению одних касательных MT, M_1T_1, M_2T_2 и т. д. (рис. 2).

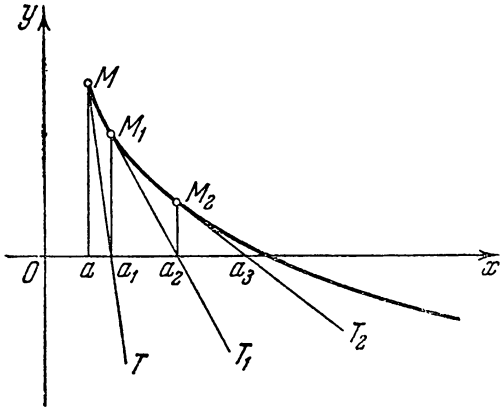

Рис. 2.

При этом по специальной формуле (см. ниже § 8) можно оценить границы погрешностей приближённых равенств: $x_1 \approx a_1$, $x_1 \approx a_2$ и т. п., но всё-таки приближение к искомому корню с двух сторон при комбинированном применении двух способов имеет за собой преимущество большей наглядности, а часто и большей простоты в практическом отношении, так как оценка погрешности при этом получается автоматически. Перейдём к примерам.

Пример 1. Найти все положительные корни уравнения

$$f(x) = 2^x - 4x = 0.$$

Из таблицы

x	0	1	2	3	4	5	6	$x > 6$
$f(x)$	1	−2	−4	−4	0	12	40	$f(x) > 0$

делаем заключение, что $0 < x_1 < 1$ и $x_2 = 4$.

Так как $f'(x) = 2^x \ln 2 - 4 \approx 0{,}693 \cdot 2^x - 4$ и $f''(x) \approx 0{,}693^2 \cdot 2^x$, то в интервале $(0, 1)$ условие (2) соблюдается:

$$f''(x) > 0,$$

а так как $f(0)>0$, а $f(1)<0$, то буквой a обозначим левый конец интервала: $a=0$; $b=1$. Тогда

$$f(a)=f(0)=1,$$
$$f(b)=f(1)=-2,$$
$$f'(a)=f'(0)\approx -3{,}307.$$

Следовательно, применяя формулы (4) и (8), получим:

$$a_1=0-\frac{1}{-3{,}307}=0{,}3023\ \ldots\ \approx 0{,}302,$$
$$b_1=0-\frac{1\,(1-0)}{-2-1}=0{,}3333\ \ldots\ \approx 0{,}334.$$

Отсюда

$$x_1\approx\frac{a_1+b_1}{2}\approx 0{,}318\ (\pm 0{,}016).$$

Примечание. Рекомендуется всегда округлять нижнюю границу интервала в сторону уменьшения, а верхнюю — в сторону увеличения (например, 0,3333 ... $=0{,}334$, а не 0,333), так как иначе можно получить такой интервал, что искомый корень окажется лежащим вне его.

Пример 2. Найти положительный корень уравнения

$$f(x)=x+4\sin x-5e^x+6=0\qquad (e\approx 2{,}7183).$$

Для определения интервала, в котором заключён корень, вычислим несколько значений функции:

$$f(0)=1,$$
$$f(1)=1+4\sin 1-5e+6\approx 7+4\sin 57^\circ 18'-5\cdot 2{,}7183\approx$$
$$\approx 7+4\cdot 0{,}8415-13{,}5915=-3{,}2255,$$
$$f(0{,}5)=0{,}5+4\sin 28^\circ 39'-5\cdot 1{,}6487+6\approx 0{,}1741.$$

Так как $f(0{,}5)>0$, а $f(1)<0$, то $0{,}5<x_1<1$; кроме того, $f''(x)=-4\sin x-5e^x<0$ в интервале (0,5; 1). Следовательно, примем $a=1$, $b=0{,}5$.

По формулам (8) и (4) получим:

$$a_1=1-\frac{-3{,}2255}{-10{,}431}=0{,}690\ \ldots\ \approx 0{,}70$$

(так как $f'(1)=1+4\cos 1-5e\approx -10{,}431$),

$$b_1=1-\frac{-3{,}2255\,(0{,}5-1)}{0{,}1741-(-3{,}2255)}=0{,}525\ \ldots\ \approx 0{,}52.$$

В этом случае получился довольно большой интервал (0,52; 0,70), так что лучше было бы предварительно провести небольшую дополни-

тельную работу (табличным способом), чтобы исходный интервал (a, b) был бы меньших размеров.

В том случае, когда в интервале (a, b) нарушается условие (2), указанный путь сближения границ интервала может оказаться неприемлемым. Действительно, если, например, кривая $y=f(x)$ будет иметь вид, указанный на рис. 3, то ни к одному из концов интервала способ Ньютона нельзя применить, так как это нас удалит от искомого корня; причиной является обращение в нуль $f''(x)$ внутри интервала (a, b); в этом случае кривая $y=f(x)$ имеет так называемую *точку перегиба*, в которой изменяется направление её вогнутости.

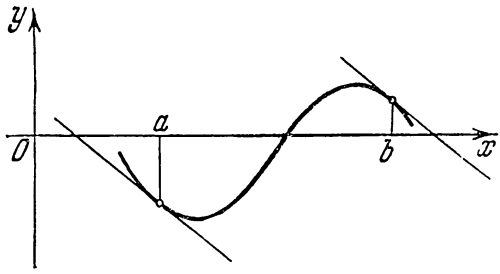

Рис. 3.

Правда, уменьшив табличным путём интервал (a, b), можно, вообще говоря, добиться выполнения условия (2) (если $f''(x_1)\neq 0$), но при $f''(x_1)=0$, естественно, для любого интервала, заключающего корень x_1 уравнения (1), условие (2) будет нарушаться.

Условие (2) является достаточным для совместного применения способов Ньютона и линейного интерполирования к интервалу (a, b); однако оно не является необходимым.

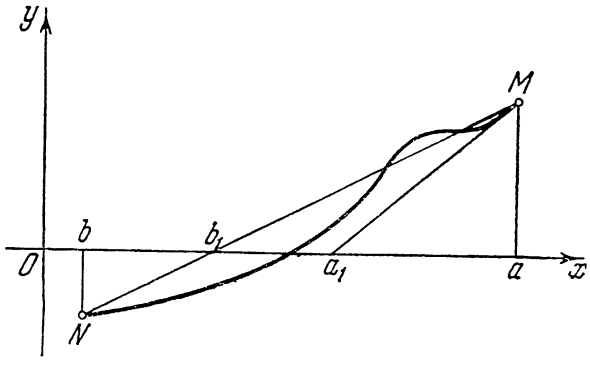

Рис. 4.

Как показывает рис. 4, иногда даже при нарушении условия (2) применение этих двух способов может дать хороший результат.

В скрытой форме способом линейного интерполирования пользуются обычно при вычислении корней табличным способом. Действительно, если, например, для уравнения $x^3+2x-28=0$ из таблицы

x	2	3
$f(x)$	−16	5

заключаем, что корень лежит в интервале (2, 3) и мы хотим уточнить цифру десятых долей искомого корня, то было бы нерационально

придавать аргументу значения: 2,1; 2,2; ..., так как, повидимому, функция $f(x)$ обращается в нуль где-то вблизи точки $x=3$.

Естественно поэтому попробовать придать аргументу значения 2,7; 2,8 и, может быть, соседние с ними.

Подтверждением этого предположения послужило бы вычисление b_1 по способу линейного интерполирования ($b_1 \approx 2{,}76$). Заметим, наконец, что, применяя табличный способ к трансцендентным уравнениям, надо быть особенно внимательным в том случае, когда, например, после убывания до малых положительных значений функция начнёт возрастать. Может быть, при этом кривая в какой-то момент коснётся Ox или даже даст две близкие точки пересечения. При небрежном применении табличного способа такие корни могут оказаться незамеченными.

Если на концах интервала (a, b), заключающего один корень уравнения (1), $f(a)$ и $f(b)$ одинаковых знаков, то можно применить либо одностороннее приближение к корню по способу Ньютона, либо один из способов, описанных в следующих параграфах.

§ 8. Обобщение способа Ньютона

Рассмотрим другой вывод формулы (8) предыдущего параграфа, что даст нам, во-первых, возможность оценить границу погрешности приближённого равенства $x_1 \approx a - \frac{f(a)}{f'(a)}$ и, во-вторых, приведёт нас к ряду формул, являющихся естественным обобщением способа Ньютона.

Пусть a — приближённое значение искомого корня уравнения

$$f(x)=0 \tag{1}$$

(алгебраического или трансцендентного), причём

$$x_1 = a + \alpha \approx a, \tag{2}$$

т. е. α есть погрешность приближённого равенства $x_1 \approx a$.

Так как $f(a+\alpha)=0$ ($a+\alpha$ — точное значение корня), то наряду с равенством (2) можно написать следующее равенство, справедливое при любом значении постоянного c:

$$x_1 = a+\alpha+cf(a+\alpha) = a+\alpha+c\left[f(a)+f'(a)\alpha+\frac{f''(\xi)}{2!}\alpha^2\right] =$$

$$= a+cf(a)+\alpha\left[1+cf'(a)\right]+c\frac{f''(\xi)}{2!}\alpha^2, \tag{3}$$

где ξ — некоторое число, заключённое между a и x_1.

Если взять

$$c = -\frac{1}{f'(a)}, \tag{4}$$

то коэффициент при α в равенстве (3) превратится в нуль, и мы будем иметь:

$$x_1 = a - \frac{f(a)}{f'(a)} - \frac{f''(\xi)}{2f'(a)}\alpha^2 = a_1 + \alpha_1 \approx a_1, \tag{3'}$$

где $a_1 = a - \frac{f(a)}{f'(a)}$ — новое (ньютоновское) приближение к искомому корню, а

$$\alpha_1 = -\frac{f''(\xi)}{2f'(a)}\alpha^2 \tag{5}$$

есть погрешность приближённого равенства $x_1 \approx a_1$.

Для получения формулы, обобщающей формулу Ньютона, напишем равенство (2) в виде

$$x_1 = a + \alpha + (c_1 + c_2\alpha) f(a+\alpha) =$$
$$= a + \alpha + (c_1 + c_2\alpha)\left[f(a) + f'(a)\alpha + \frac{f''(a)}{2!}\alpha^2 + \frac{f'''(\xi)}{3!}\alpha^3\right] \tag{6}$$

или, вводя обозначения

$$f(a) = F, \quad f'(a) = F', \quad \frac{f''(a)}{2!} = F'', \quad \frac{f'''(\xi)}{3!} = F'''(\xi), \tag{7}$$

перепишем равенство (6) в виде

$$x_1 = a + c_1F + \alpha\left[1 + c_1F' + c_2F\right] + \alpha^2\left[c_1F'' + c_2F'\right] +$$
$$+ \alpha^3\left[c_1F'''(\xi) + c_2F''\right] + \alpha^4 c_2F'''(\xi). \tag{6'}$$

Последнее соотношение справедливо при любых постоянных c_1 и c_2; подберём их так, чтобы обратились в нуль коэффициенты при α и α^2 в правой части равенства (6'):

$$1 + c_1F' + c_2F = 0, \quad c_1F'' + c_2F' = 0. \tag{8}$$

Из уравнений (8) получим:

$$c_1 = \frac{-F'}{(F')^2 - FF''}, \quad c_2 = \frac{F''}{(F')^2 - FF''}. \tag{9}$$

Подставляя полученные значения c_1 и c_2 в (6'), получим:

$$x_1 = a - \frac{FF'}{(F')^2 - FF''} + \frac{\alpha^3\left[(F'')^2 - F'F'''(\xi)\right] + \alpha^4F''F'''(\xi)}{(F')^2 - FF''} \tag{10}$$

или

$$x_1 = a_1^* + \alpha_1^* \approx a_1^*, \tag{10'}$$

где

$$a_1^* = a - \frac{FF'}{(F')^2 - FF''} = a - \frac{2f(a)f'(a)}{2\left[f'(a)\right]^2 - f(a)f''(a)} \tag{11}$$

и

$$\alpha_1^* = \frac{\alpha^3\left[(F'')^2 - F'F'''(\xi)\right] + \alpha^4F''F'''(\xi)}{(F')^2 - FF''}. \tag{12}$$

Если бы мы взяли множитель при $f(a+\alpha)$ в равенстве типа (6) в виде $c_1+c_2\alpha+c_3\alpha^2$, то смогли бы получить:

$$x_1=a_1^{**}+\alpha_1^{**}\approx a_1^{**}, \tag{13}$$

где

$$a_1^{**}=a-\frac{F\begin{vmatrix}F' & F\\ F'' & F'\end{vmatrix}}{\begin{vmatrix}F' & F & 0\\ F'' & F' & F\\ F''' & F'' & F'\end{vmatrix}}, \tag{14}$$

а α_1^{**} зависит от α^4, α^5, α^6.

Формулы (11) и (14) обобщают способ Ньютона и являются частными случаями полученной автором настоящей статьи общей формулы, которая даёт возможность получить в один приём вместо исходного приближения $x_1=a+\alpha\approx a$ новое приближение $x_1\approx\bar{a}_1$, погрешность которого зависит от α^{k+1}, α^{k+2}, ... (где k — любое целое число).

Формулы (5) и (12) можно применить для оценки верхней границы погрешности приближённых значений a_1 и a_1^* искомого корня: зная, что $|\alpha|<\gamma$ (где γ равна 0,1 или 0,01 и т. п.), определяют сначала границы, в которых заключена $f''(\xi)$ или $f'''(\xi)$, а затем ищут по формулам (5) или (12) верхнюю границу для $|\alpha_1|$ или для $|\alpha_1^*|$.

Пример. Вычислить положительный корень уравнения

$$f(x)=x^3+3x-3=0,$$

найдя предварительно исходное приближение по способу Горнера с точностью до 0,01.

Так как $0<x_1<1$, то, уточнив число десятых, найдём:

$$0{,}8<x_1<0{,}9.$$

Делаем подстановку $x-0{,}8=y$ (отказавшись в дальнейшем от подстановок $y=\frac{Y}{10}$, $z=\frac{Z}{10}$ и т. п.):

	1	0	3	−3
0,8	1	0,8	3,64	−0,088
0,8	1	1,6	4,92	
0,8	1	2,4		

$$f(x)=(x-0{,}8)^3+2{,}4(x-0{,}8)^2+4{,}92(x-0{,}8)-0{,}088=$$
$$=y^3+2{,}4y^2+4{,}92y-0{,}088=\varphi(y)=0. \tag{15}$$

Так как $0{,}01 < y_1 < 0{,}02$, то делаем подстановку $y - 0{,}01 = z$ (следовательно, $z = x - 0{,}81$)

	1	2,4	4,92	−0,088
0,01	1	2,41	4,9141	−0,038559
0,01	1	2,42	4,9383	
0,01	1	2,43		

$$f(x) = (y - 0{,}01)^3 + 2{,}43\,(y - 0{,}01)^2 + 4{,}9683\,(y - 0{,}01) - 0{,}038559 =$$
$$= z^3 + 2{,}43z^2 + 4{,}9683z - 0{,}038559,$$

а так как $z = x - 0{,}81$, то коэффициенты последнего многочлена равны тейлоровским коэффициентам разложения $f(x)$ по степеням $x - 0{,}81$, т. е.

$$f\,(0{,}81) = -0{,}038559 = F,$$
$$f'\,(0{,}81) = +4{,}9683 = F',$$
$$\frac{f''\,(0{,}81)}{2!} = 2{,}43 = F'',$$
$$\frac{f'''\,(0{,}81)}{3!} = 1 = F'''.$$

Кроме того, $0{,}81 < x_1 < 0{,}82$, т. е. $x_1 = 0{,}81 + \alpha \approx 0{,}81 = a$, где $0 < \alpha < 0{,}01$.

Так как $0{,}81 < \xi < x_1 < 0{,}82$, $f''(x) = 6x$, $f'''(x) = 6$, то $f''(\xi) < 4{,}92$ и $f'''(\xi) = 6$ или $F'''(\xi) = 1$. Формула (5) даёт нам:

$$\alpha_1 < 0 \text{ и } |\alpha_1| < \frac{4{,}92}{2 \cdot 4{,}9683} \cdot |\alpha|^2 < 0{,}00005,$$

а по формуле (12) имеем:

$$0 < \alpha_1^* = \frac{\alpha^3\,[2{,}43^2 - 4{,}9683 \cdot 1] + \alpha^4 \cdot 2{,}43 \cdot 1}{4{,}9683^2 - 2{,}43 \cdot (-0{,}038559)} < \frac{\alpha^3 + 3\alpha^4}{24} < 0{,}00000005,$$

или

$$0 < \alpha_1^* < 0{,}00000005. \qquad (16)$$

Следовательно, вычисление a_1 по способу Ньютона будем вести до 5-го, а вычисление a_1^* по формуле (11) — до 8-го знака после запятой:

$$a_1 = 0{,}81 - \frac{-0{,}038559}{4{,}9683} = 0{,}81776\ \ldots,$$

$$a_1^* = 0{,}81 - \frac{-0{,}038559 \cdot 4{,}9683}{(4{,}9683)^2 - 2{,}43 \cdot (-0{,}038559)} = 0{,}81773165\ \ldots$$

Итак, по способу Ньютона

$$0{,}81771 < x_1 < 0{,}81777,$$

а по обобщённому способу

$$0{,}81773165 < x_1 < 0{,}81773171. \tag{17}$$

Примечание 1. Из (17) видно, что для погрешности равенства $x_1 \approx 0{,}81$ имеем: $\alpha < 0{,}0078$; следовательно, вместо (16) можно получить: $\alpha_1^* < 2 \cdot 10^{-8}$, а вместо (17):

$$0{,}81773165 < x_1 < 0{,}81773168. \tag{17'}$$

Примечание 2. Так как мы применяли способ Ньютона не в комбинации со способом линейного интерполирования и погрешность α исходного приближения мала, то можно было отказаться от требования равенства знаков у $f(a)$ и $f''(a)$. Правда, в данном случае лучше было бы взять $a \approx 0{,}82$, так как это значение ближе к искомому корню, как это видно из уравнения (15). Рекомендуем читателю проделать соответствующие вычисления; они должны дать для x_1 лучшие границы, чем (17'), если, конечно, учесть после первого этапа вычислений, что в этом случае

$$-0{,}0023 < \alpha < 0.$$

Конечно, обобщённые формулы типа (14) слишком громоздки, чтобы их применять к более или менее сложным уравнениям; особенно много хлопот доставляет оценка погрешности. Однако для простейших вопросов, с которыми приходится сталкиваться в средней школе, например для извлечения корней, дело сильно упрощается, и обобщённый способ Ньютона даёт, как увидим в следующем параграфе, ряд полезных формул.

Отметим, наконец, что для алгебраических уравнений определённой степени с буквенными коэффициентами можно совершенно элементарно вывести формулы, отвечающие обычному или обобщённому способу Ньютона, стоит только разлагать $f(a+\alpha)$ не по формуле Тейлора, а элементарным путём; в § 10 это будет проделано для уравнения $x^n - N = 0$.

§ 9. Способ итерации

Для применения способа итерации, получившего своё название от латинского слова iteratio — повторение, требуется предварительное преобразование данного уравнения

$$f(x) = 0 \tag{1}$$

к виду

$$x = \varphi(x). \tag{2}$$

Как увидим ниже, такое преобразование можно осуществить многими способами.

Допустим, что для интересующего нас действительного корня x_1 уравнения (2) мы получили каким-нибудь способом (например, табличным) приближённое значение $x_1 \approx a$. Вычисляя затем последовательно ряд чисел $a_1, a_2, \ldots, a_n, \ldots$ по формуле

$$a_k = \varphi(a_{k-1}) \tag{3}$$

[считая $a_0 = a$, следовательно, $a_1 = \varphi(a)$], мы при некоторых условиях сможем получить значение корня x_1 с какой угодно малой погрешностью. Следующая теорема даёт условие, обеспечивающее сходимость чисел $a_1, a_2, a_3, \ldots$ к искомому корню.

Теорема. *Если в интервале, заключающем числа x_1, а и числа $a_1, a_2, \ldots$, вычисляемые по формуле* (3), *выполнено условие*

$$|\varphi'(x)| < m < 1, \tag{4}$$

то

$$\lim_{n\to\infty} a_n = x_1.$$

Другими словами, повторяя одну и ту же операцию, определяемую равенством (3), мы сможем приблизиться к искомому корню как угодно близко. Для доказательства теоремы из равенства

$$x_1 = \varphi(x_1) \tag{5}$$

[x_1 — корень уравнения (2)] вычтем почленно равенство (3) и применим теорему о среднем значении:

$$x_1 - a_k = \varphi(x_1) - \varphi(a_{k-1}) = (x_1 - a_{k-1})\varphi'(\xi),$$

где ξ — некоторое число, лежащее в интервале (a_{k-1}, x).

Переходя к абсолютным значениям и используя условие (4), получим:

$$|x_1 - a_k| = |x_1 - a_{k-1}| \cdot |\varphi'(\xi)| < m|x_1 - a_{k-1}|. \tag{6}$$

По условию теоремы неравенство (6) имеет место для $k = 0, 1, 2, 3 \ldots$; следовательно, если обозначим для краткости $|x_1 - a_s| = \Delta_s$, то будем иметь:

$$\left.\begin{aligned} \Delta_1 &< m\Delta_0, \\ \Delta_2 &< m\Delta_1, \\ &\cdots\cdots\cdots \\ \Delta_n &< m\Delta_{n-1}, \end{aligned}\right\} \tag{7}$$

откуда

$$\Delta_1\Delta_2 \ldots \Delta_{n-1}\Delta_n < m^n\Delta_0\Delta_1 \ldots \Delta_{n-1},$$

или

$$\Delta_n < m^n\Delta_0,$$

т. е.

$$|x_1 - a_n| < m^n|x_1 - a_0|.$$

Так как $\lim\limits_{n\to\infty} m^n = 0$, то

$$\lim_{n\to\infty} a_n = x_1,$$

т. е. теорема доказана.

Очевидно, числа a_1, a_2, ..., a_n тем быстрее будут приближаться к корню x_1, чем меньше будет m. Следовательно, заменяя уравнение (1) уравнением (2), надо стремиться получить справа такую функцию $\varphi(x)$, производная которой в окрестности искомого корня была бы мала по абсолютному значению.

Отметим, что при вычислении с определённым числом десятичных знаков из совпадения какого-то a_s с a_{s-1} будет следовать, что $a_s = \varphi(a_{s-1}) = a_{s-1}$, т. е. a_{s-1} будет (в пределах точности вычислений) удовлетворять уравнению (2) и можно считать, что $x_1 \approx a_{s-1}$; следовательно, процесс вычислений в способе итерации прекращается автоматически.

Пример. Решить уравнение

$$f(x) = x^3 - 20x + 4 = 0. \tag{8}$$

Как легко установить табличным способом, уравнение это имеет три действительных корня x_1, x_2, x_3, лежащих, соответственно, в интервалах (0, 1), (4, 5), (— 5, — 4).

Заменить уравнение (8) уравнением вида (2) можно многими способами; например,

$$x = \frac{x^3 + 4}{20} = 0{,}05\,(x^3 + 4), \tag{9}$$

или

$$x = \sqrt[3]{20x - 4}, \tag{10}$$

или

$$x = x + (x^3 - 20x + 4). \tag{11}$$

Однако для уравнения (11) $\varphi'(x) = 3x^2 - 19$, и условие (4) не соблюдается ни для одного из интервалов, заключающих наши корни.

Для уравнения (9) $\varphi'(x) = \frac{3x^2}{20}$ и в интервале (0, 1) $|\varphi'(x)| < \frac{3}{20}$, следовательно, x_1 можно найти, решая уравнение (9) по способу итерации.

Но в окрестностях корней x_2 и x_3 для уравнения (9) $|\varphi'(x)| > \frac{48}{20}$, следовательно, x_2 и x_3 не следует искать по способу итерации из уравнения (9).

Для уравнения же (10) $\varphi'(x) = \frac{20}{3\sqrt[3]{(20x-4)^2}}$; в интервале (0, 1) $|\varphi'(x)| > 1$, и способ итерации непригоден, в интервалах же (— 5, — 4) и (4, 5) $|\varphi'(x)| < 1$, т. е. корни x_2 и x_3 можно найти из уравнения (10).

Найдём x_1 из уравнения (9). Пусть $x_1 \approx 0 = a$. Тогда

$$a_1 = 0{,}05\,(0^3 + 4) = 0{,}2,$$
$$a_2 = 0{,}05\,(0{,}2^3 + 4) = 0{,}2004,$$
$$a_3 = 0{,}05\,(0{,}2004^3 + 4) \approx 0{,}2004.$$

Так как $a_3 = a_2$ (в пределах точности вычислений), то $x_1 \approx 0{,}2004$.

Читателю рекомендуется вычислить x_1, взяв за исходное приближение $x_1 \approx 1 = a$. При этом удобно применять «Таблицы Барлоу квадратов, кубов, корней квадратных, корней кубичных и обратных величин» издания 1933 г.

Конечно, результат должен получиться тот же: $x_1 \approx 0{,}2004$.

Найдём ещё x_3 из уравнения (10). Пусть $x_3 \approx -5 = a$. Тогда

$$a_1 = \sqrt[3]{-104} \approx -4{,}703,$$
$$a_2 = \sqrt[3]{-98{,}06} \approx -4{,}611,$$
$$a_3 = \sqrt[3]{-96{,}44} \approx -4{,}586,$$
$$a_4 = \sqrt[3]{-95{,}72} \approx -4{,}574,$$
$$a_5 = \sqrt[3]{-95{,}48} \approx -4{,}571,$$
$$a_6 = \sqrt[3]{-95{,}42} \approx -4{,}570,$$
$$a_7 = \sqrt[3]{-95{,}40} \approx -4{,}569,$$
$$a_8 = \sqrt[3]{-95{,}38} \approx -4{,}569.$$

Здесь корни извлекались по таблицам Барлоу; в столбце кубов (стр. 94 и 93) находилось число, близкое к подкоренному (при перенесении запятой на 9 цифр), а в столбце с заголовком n получалось значение корня.

Так как $a_8 \approx a_7$ (при вычислениях с тремя знаками после запятой), то вычисления автоматически обрываются, и мы имеем: $x_3 \approx -4{,}569$.

В общем случае простейший путь для замены уравнения (1) уравнением вида (2) при соблюдении условия (4) заключается в следующем: так как уравнение

$$x = x - cf(x) \tag{12}$$

при любом постоянном c имеет такие же корни, как и уравнение (1), то надо лишь подобрать постоянное c так, чтобы $\varphi'(x) = 1 - cf'(x)$ была по абсолютной величине достаточно мала в окрестности искомого корня, а так как в нашем распоряжении сначала имеется лишь

приближённое значение a искомого корня, то проще всего c найти из уравнения

$$\varphi'(a) = 1 - cf'(a) = 0.$$

Имеем: $c = -\frac{1}{f'(a)}$, и следовательно, уравнение (12) примет вид

$$x = x - \frac{f(x)}{f'(a)}. \tag{13}$$

Если a лежит достаточно близко около x_1 и $f'(x)$ меняется медленно, т. е. $f'(a_1)$, $f'(a_2)$ и т. д. мало отличаются от $f'(a)$ (другими словами, в окрестности искомого корня $f''(x)$ невелика по абсолютной величине), то $\varphi'(a_1)$, $\varphi'(a_2)$, $\varphi'(a_3)$ и т. п. будут мало отличаться от нуля, следовательно, условие (4) будет выполнено, и уравнение (13) можно будет решать по способу итерации. Будем иметь:

$$a_1 = a - \frac{f(a)}{f'(a)}, \tag{14}$$

что полностью совпадает с формулой, соответствующей способу Ньютона; следовательно, a_1 есть абсцисса точки пересечения касательной AT (рис. 5) с осью Ox.

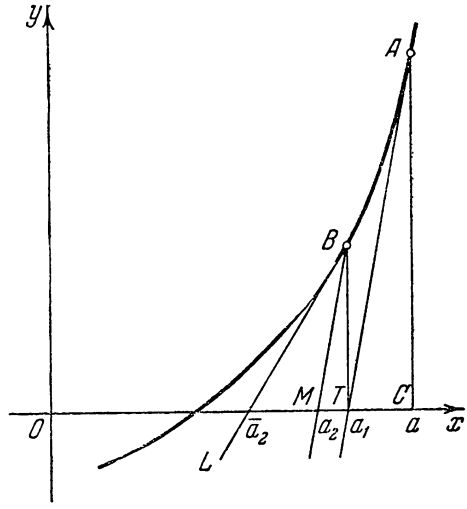

Рис. 5.

Но в способе Ньютона и следующее приближение (обозначим его через $\bar{a}_2$), получаемое по формуле

$$\bar{a}_2 = a_1 - \frac{f(a_1)}{f'(a_1)}, \tag{15}$$

соответствует касательной BL в точке B.

По способу же итерации имеем:

$$a_2 = a_1 - \frac{f(a_1)}{f'(a)}, \tag{16}$$

что связано с проведением прямой BM, параллельной AT; действительно, из рис. 5 имеем:

$$\operatorname{tg} \widehat{BMx} = \frac{BT}{MT} = \frac{f(a_1)}{a_1 - a_2},$$

$$\operatorname{tg} \widehat{ATx} = \frac{AC}{TC} = \frac{f(a)}{a - a_1},$$

а обе последние дроби в силу (14) и (16) равны $f'(a)$. Ясно, что по способу Ньютона мы быстрее будем приближаться к искомому

корню, чем при решении уравнения (13) по способу итерации, но зато в последнем случае проще вычисления, так как не надо вычислять $f'(a_1)$, $f'(a_2)$ и т. д.; кроме того, постоянный делитель $f'(a)$ можно взять с некоторым удобным для вычислений округлением.

Отметим, что способ Ньютона есть разновидность способа итерации, так как уравнение $f(x)=0$ и уравнение

$$x=x-\frac{f(x)}{f'(x)} \tag{17}$$

будут иметь одинаковые корни, если только $f'(x)$ не обращается в ∞ ни при каких значениях аргумента; решение же уравнения (17) по способу итерации равносильно решению уравнения (1) по способу Ньютона.

§ 10. Различные способы извлечения корней из чисел

Во введении указывалось на желательность ознакомления учащихся с табличным способом решения уравнений. Однако метод этот связан с большой вычислительной работой, и естественно, у учащихся может возникнуть желание познакомиться с более совершенными приёмами решения уравнений.

Многие из способов, изложенных в предыдущих параграфах, доступны пониманию учащихся, если излагать их не в общем виде, а на конкретных примерах уравнений невысоких степеней с буквенными коэффициентами, где можно обойтись без формулы Тейлора, без понятия производной функции.

Ниже даётся применение некоторых из этих методов к решению двучленных уравнений, т. е. к извлечению корней из чисел.

Конечно, лучше всего знакомить учащихся с этими способами в порядке кружковой работы, хотя опытный преподаватель, может быть, найдёт уместным кое-что изложить и в классе.

Извлечение квадратных корней по способу Лагранжа. Найдём для примера положительный корень x_1 уравнения

$$x^2-2x-2=0, \tag{1}$$

$$2<x_1<3,\quad x=2+\frac{1}{y};$$

	1	−2	−2
2	1	0	−2
2	1	2	

следовательно, $2y^2-2y-1=0$, $1<y_1<2$, $y=1+\frac{1}{z}$:

	2	−2	−1
1	2	0	−1
1	2	2	

следовательно, $z^2 - 2z - 2 = 0$; так как коэффициенты этого уравнения полностью совпадают с коэффициентами уравнения (1), то в дальнейшем коэффициенты получающихся уравнений и их корни будут периодически повторяться:

$$z_1 = 2 + \frac{1}{u_1}, \quad u_1 = 1 + \frac{1}{v_1}, \quad v_1 = 2 + \frac{1}{w_1}, \ldots,$$

следовательно,

$$x_1 = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \ddots}}} = [2,\ 1,\ 2,\ 1,\ 2, \ldots]$$

(в скобках написаны неполные частные непрерывной дроби).

В теории чисел доказывается (см., например, Арнольд «Теория чисел»), что при разложении в непрерывную дробь иррациональных чисел вида $A + B\sqrt{C}$, где A, B, C — рациональные числа, получающаяся непрерывная дробь будет обязательно бесконечной периодической дробью.

Следовательно, действительные корни квадратных уравнений и, в частности $\sqrt{N}$, если N — неточный квадрат, будут выражаться бесконечными периодическими дробями.

Рекомендуем читателю убедиться, что, например,

$$\sqrt{2} = [1,\ 2,\ 2,\ 2,\ 2, \ldots],$$
$$\sqrt{3} = [1,\ 1,\ 2,\ 1,\ 2, \ldots],$$
$$\sqrt{28} = [5,\ 3,\ 2,\ 3,\ 10,\ 3,\ 2,\ 3,\ 10, \ldots].$$

(В конце книги Арнольда дана таблица разложений $\sqrt{N}$ для всех целых чисел от 2 до 99.)

Имея же в своём распоряжении для $\sqrt{N}$ неограниченный ряд неполных частных, можно, составляя подходящие дроби, найти просто и быстро $\sqrt{N}$ с любой точностью. Найдём, например, $\sqrt{28}$ с точностью до 10^{-6}:

P_k	5	16	37	127	1307	4048	
Q_k	1	3	7	24	247	765	1777
q_k	5	3	2	3	10	3	2 ...

Так как $\frac{1}{765 \cdot 1777} < \frac{1}{1\,300\,000} < 0{,}0000008$, то

$$\sqrt{28} \approx \frac{4048}{765} = 5{,}29150326 \ldots \text{ (с избытком)},$$

причём абсолютная величина ошибки меньше 0,0000008.

Извлечение корней по способу Горнера. Как известно, изучаемый в школе алгорифм извлечения корней квадратных из чисел просто обобщается на случай извлечения корней третьей степени. Схематично процесс можно представить так:

$$\begin{array}{l} \sqrt[3]{N} = c_0, c_1 \ldots \\ - \quad \overline{c_0^3} \\ \hline N - c_0^3 \\ - \quad 3c_0^2\left(\frac{c_1}{10}\right) + 3c_0\left(\frac{c_1}{10}\right)^2 + \left(\frac{c_1}{10}\right)^3 \\ \hline N - \left(c_0 + \frac{c_1}{10}\right)^3 \quad \text{и т. д.} \end{array}$$

Пусть целая часть корня c_0 уже найдена, и мы ищем цифру десятых долей, т. е. наибольшую цифру c_1, для которой сумма трёх слагаемых

$$3c_0^2\left(\frac{c_1}{10}\right) + 3c_0\left(\frac{c_1}{10}\right)^2 + \left(\frac{c_1}{10}\right)^3$$

остаётся ещё меньше, чем $N - c_0^3$.

Найдя c_1, а затем и $N - \left(c_0 + \frac{c_1}{10}\right)^3$, ищем цифру сотых долей и т. д. Но то же самое делается и в способе Горнера, если представить соответствующее уравнение в виде $-x^3 + N = 0$: найдя c_0, делаем подстановку $x - c_0 = y$:

	-1	0	0	N
c_0	-1	$-c_0$	$-c_0^2$	$N - c_0^3$
c_0	-1	$-2c_0$	$-3c_0^2$	
c_0	-1	$-3c_0$		

а найдя c_1, — подстановку $y - \frac{c_1}{10} = z$:

	-1	$-3c_0$	$-3c_0^2$	$N - c_0^3$
$\frac{c_1}{10}$	-1	$-3c_0 - \frac{c_1}{10}$	$-3c_0^2 - 3c_0\frac{c_1}{10} - \left(\frac{c_1}{10}\right)^2$	$N - \left(c_0 + \frac{c_1}{10}\right)^3$
$\frac{c_1}{10}$	-1	$-3c_0 - \frac{2c_1}{10}$	$-3\left(c_0 + \frac{c_1}{10}\right)^2$	
$\frac{c_1}{10}$	-1	$-3\left(c_0 + \frac{c_1}{10}\right)$		

(обходясь без подстановок $y = \frac{Y}{10}$, $z = \frac{Z}{10}$ и т. п.). В результате

получаем: $N-\left(c_0+\frac{c_1}{10}\right)^3$ и коэффициенты нового уравнения, которые при подстановке $z-\frac{c_2}{100}=u$ дадут $N-\left(c_0+\frac{c_1}{10}+\frac{c_2}{100}\right)^3$ и т. д.

Здесь бросается в глаза черта, характерная вообще для способа Горнера: вычисляя $f\left(c_0+\frac{c_1}{10}\right)$, мы используем найденное уже значение $f(c_0)$ и лишь несколько «подправляем» его; затем, подправив немного $f\left(c_0+\frac{c_1}{10}\right)$, получаем: $f\left(c_0+\frac{c_1}{10}+\frac{c_2}{100}\right)$ и т. д. В этом отношении способ Горнера выгодно отличается от табличного способа, где значения функции вычисляются независимо друг от друга.

Очевидно, при вычислении $\sqrt[n]{N}$ дело будет обстоять так же, как и при извлечении корня третьей степени.

Формулы, вытекающие из обычного и обобщённого способов Ньютона. Используя идею способа Ньютона и его обобщений (см. § 8), можно элементарным путём получить ряд практически удобных формул, пригодных для извлечения корней любой степени из чисел. Пусть $x_1=\sqrt[n]{N}$ — корень уравнения

$$x^n-N=0, \tag{2}$$

причём первое приближение его известно:

$$x_1=a+\alpha\approx a.$$

Так как $(a+\alpha)^n-N=0$, то имеем:

$$x_1=a+\alpha+(c_1+c_2\alpha+c_3\alpha^2+\ldots+c_k\alpha^{k-1})\ [(a+\alpha)^n-N]=$$
$$=a+\alpha+(c_1+c_2\alpha+\ldots+c_k\alpha^{k-1})\ [a^n-N+\tbinom{n}{1}a^{n-1}\alpha+$$
$$+\tbinom{n}{2}a^{n-2}\alpha^2+\tbinom{n}{3}a^{n-3}\alpha^3+\ldots]\ ^1). \tag{3}$$

Взяв

$$c_1=-\frac{1}{na^{n-1}},\ \text{а}\quad c_2=c_3=\ldots=c_k=0,$$

получим:

$$x_1=a-\frac{a^n-N}{na^{n-1}}+\alpha_1\approx a-\frac{a^n-N}{na^{n-1}}=a_1, \tag{4}$$

где

$$\alpha_1\approx-\frac{n-1}{2a}\alpha^2$$

(если пренебречь членами, содержащими α^3, α^4, ...).

Но можно подобрать в формуле (3) числа c_1, c_2, ..., c_k так, чтобы в правой части исчезли члены, содержащие α, α^2, ..., α^k;

---

[1]) Здесь биномиальные коэффициенты C_n^1, C_n^2 и т. д. обозначены через $\binom{n}{1}$, $\binom{n}{2}$ и т. д.

тогда получим формулу, дающую приближение, погрешность которого будет зависеть от $(k+1)$-й и более высоких степеней α.

Нетрудно убедиться, что, например, для $\sqrt[3]{N}$, если исходное приближение $\sqrt[3]{N}=a+\alpha\approx a$, то

	следующее приближение		его погрешность	
при $k=1$	$a_1=\frac{2a^3+N}{3a^2}=a-\frac{a^3-N}{3a^2}$,	(5)	$\alpha_1=-\frac{3a\alpha^2+\alpha^3}{3a^2}\approx-\frac{\alpha^2}{a}$,	(5′)
при $k=2$	$a_1^*=a-\frac{(a^3-N)a}{2a^3+N}=\frac{a(a^3+2N)}{2a^3+N}$,	(6)	$\alpha_1^*=\frac{2a\alpha^3+\alpha^4}{2a^3+N}\approx\frac{2\alpha^3}{3a^2}$,	(6′)
при $k=3$	$a_1^{**}=a-\frac{(a^3-N)(6a^4+3aN)}{10a^6+16a^3N+N^2}$,	(7)	$\alpha_1^{**}\approx-\frac{\alpha^4}{3a^3}$.	(7′)

Если же $\sqrt[5]{N}=a+\alpha\approx a$, то

	следующее приближение		его погрешность	
при $k=1$	$a_1=a-\frac{a^5-N}{5a^4}$,	(8)	$\alpha_1\approx-\frac{2\alpha^2}{a}$,	(8′)
при $k=2$	$a_1^*=\frac{a(2a^5+3N)}{3a^5+2N}$,	(9)	$\alpha_1^*\approx\frac{2\alpha^3}{a^2}$.	(9′)

Для корня квадратного можно получить особенно простым способом общую формулу.

Пусть $\sqrt{N}=a+\alpha$; тогда $(-\alpha)^m=(a-\sqrt{N})^m$ или

$$(-\alpha)^m=a^m-\binom{m}{1}a^{m-1}\sqrt{N}+\binom{m}{2}a^{m-2}N-\binom{m}{3}a^{m-3}N\sqrt{N}+\ldots,$$

откуда

$$\sqrt{N}=\frac{a^m+\binom{m}{2}a^{m-2}N+\binom{m}{4}a^{m-4}N^2+\ldots}{\binom{m}{1}a^{m-1}+\binom{m}{3}a^{m-3}N+\binom{m}{5}a^{m-5}N^2+\ldots}+\alpha_1, \qquad (10)$$

где

$$\alpha_1=\frac{(-1)^{m+1}\alpha^m}{\binom{m}{1}a^{m-1}+\binom{m}{3}a^{m-3}N+\ldots}\approx\frac{(-1)^{m+1}\alpha^m}{2^{m-1}a^{m-1}}. \qquad (11)$$

В частности получим:

	новое приближение		его погрешность	
при $m=2$	$a_1=\frac{a^2+N}{2a}$,	(12)	$\alpha_1=-\frac{\alpha^2}{2a}$,	(12′)
при $m=3$	$a_1^*=\frac{a^3+3aN}{3a^2+N}$,	(13)	$\alpha_1^*=\frac{\alpha^3}{3a^2+N}\approx\frac{\alpha^3}{4a^2}$,	(13′)
при $m=4$	$a_1^{**}=\frac{a^4+6a^2N+N^2}{4a^3+4aN}$,	(14)	$\alpha_1^{**}=-\frac{\alpha^4}{4a^3+4aN}\approx-\frac{\alpha^4}{8a^3}$,	(14′)
при $m=5$	$a_1^{***}=\frac{a^5+10a^3N+5aN^2}{5a^4+10a^2N+N^2}$,	(15)	$\alpha_1^{***}=\frac{\alpha^5}{5a^4+10a^2N+N^2}\approx\frac{\alpha^5}{16a^4}$.	(15′)

Приведём ещё формулу для $\sqrt[n]{N}$ при $k=2$. Если $\sqrt[n]{N}=a+\alpha\approx a$, то согласно формулам (10'), (11) и (12) из § 8

$$\sqrt[n]{N}=a+\alpha_1^*+\alpha_1^*\approx a_1^*=\frac{a[(n-1)a^n+(n+1)N]}{(n+1)a^n+(n-1)N}, \qquad (16)$$

где

$$\alpha_1^*\approx\frac{\alpha^3(n-1)(n+1)}{12a^2}. \qquad (17)$$

Рассмотрим несколько пояснительных примеров.

Пример 1. Пусть $\sqrt{13}=4+\alpha\approx 4=a$. По формуле (15) получим:

$$\sqrt{13}=a_1^{***}+\alpha_1^{***}\approx a_1^{***}=\frac{4^5+10\cdot 4^3\cdot 13+5\cdot 4\cdot 13^2}{5\cdot 4^4+10\cdot 4^2\cdot 13+13^2}=$$

$$=\frac{12\,724}{3529}=3{,}605553981\ldots$$

Здесь $\alpha_1^{***}=\frac{\alpha^5}{3529}$ (точно), а так как $|\alpha|<1$, то в полученном числе первые три знака после запятой верные, следовательно, $-0{,}4<\alpha<0$, но тогда $\alpha^{***}<0$, причём

$$\left|\alpha_1^{***}\right|<\frac{0{,}4^5}{3529}<0{,}00000291.$$

Таким образом, $\sqrt{13}\approx 3{,}605554$ (с избытком), причём абсолютная величина ошибки меньше 0,000003.

Покажем на этом примере одну интересную особенность, присущую обобщённому способу Ньютона, когда он применяется к алгебраическим уравнениям, т. е. когда можно установить простую зависимость между новой и старой погрешностями.

Мы получили $\sqrt{13}$ с пятью совершенно надёжными цифрами после запятой $\sqrt{13}=3{,}60555$. Следовательно, и $\alpha\approx -039445$, — с пятью верными цифрами после запятой; но тогда

$$\alpha_1^{***}\approx\frac{(-0{,}39445)^5}{3529}\approx -0{,}000002706$$

тоже по крайней мере с четырьмя верными значащими цифрами. Если теперь прибавим к полученному выше значению a_1^{***} эту отрицательную поправку, то получим:

$$\sqrt{13}\approx 3{,}605551275,$$

где после запятой все знаки верные, кроме, может быть, последнего.

Пример 2. $\sqrt[3]{29} = 3 + \alpha \approx 3 = a$; так как $3{,}1^3 > 29$, то $0 < \alpha < 0{,}1$. Беря формулу (6), получим:

$$a_1^* = \frac{3 \cdot (27 + 58)}{54 + 29} = \frac{255}{83} = 3{,}072288\ldots,$$

причём

$$\alpha_1^* \approx \frac{2\alpha^3}{3a^2} < 0{,}0001, \quad \text{т. е. } \alpha \approx 0{,}072;$$

следовательно, $\alpha_1^* \approx \frac{2 \cdot 0{,}072^3}{3 \cdot 3^2} \approx 0{,}00003$, т. е. $\sqrt[3]{29} \approx 3{,}07232$, где несколько сомнительна лишь последняя цифра.

О способе итерации. Любая из формул типа (5), (6), (12) и т. п. может быть применена многократно; например, найдя по формуле (6) a_1^*, можно вычислить $a_2^* = a_1^* - \frac{(a_1^{*3} - N)a_1^*}{2a_1^{*3} + N}$, затем a_3^* и т. д. Но это значит, что мы решаем по способу итерации уравнение

$$x = x - \frac{(x^3 - N)x}{2x^3 + N} = \varphi(x), \tag{18}$$

имеющее те же корни, что и уравнение $x^3 - N = 0$ (лишний корень $x = 0$ ничему не мешает).

Примечание. При этом очень быстрая сходимость процесса вытекает из того, что

$$\varphi'(x) = \frac{2(x^3 - N)^2}{(3x^3 + N)^2},$$

а следовательно, если погрешность α исходного приближения $\sqrt[3]{N} \approx a$ мала, то $|\varphi'(x)|$ будет очень мала в окрестности чисел a_1^*, a_2^* и т. д.

Отметим, что формула (12) была использована Ньютоном для итеративного улучшения равенства $\sqrt{N} \approx a$ [1]).

При таком освещении вопроса об извлечении корней становится понятной возможность решения по способу итерации любых уравнений, если они предварительно переписаны в подходящей форме.

Задачи к главе II

10. Вычислить по способу итерации с точностью до 10^{-3} иррациональный корень уравнения $x = \frac{2^x}{4}$, приняв $x \approx 0{,}5 = a$.

11. Вычислить по способу Ньютона с точностью до 0,001 корни уравнений:

а) $x - \cos x = 0$, б) $x - 2\sin x = 0$,

определив исходные приближения табличным способом с точностью до 0,1.

[1]) Интересные подробности об этом читатель найдёт в книге: Уиттекер и Робинсон, Математическая обработка результатов наблюдений, 1933, стр. 77—80.

12. Разложив в непрерывные дроби положительные корни уравнений:

$$\text{а) } x^2-4x-3=0, \quad \text{б) } 3x^2+2x-5=0, \quad \text{в) } x^2-23=0,$$

вычислить их с точностью до 10^{-6}.

13. Применяя формулу (15) из § 10 к приближённым равенствам $\sqrt{50}\approx 7$, $\sqrt{33}\approx 6$, $\sqrt{97}\approx 10$, найти для каждого случая следующее приближение и оценить его погрешность.

14. Вывести элементарным путём формулу (соответствующую способу Ньютона)

$$a_1 = a - \frac{a^3+Aa^2+Ba+C}{3a^2+2Aa+B}$$

для улучшения приближённого значения корня $x\approx a$ уравнения

$$x^3+Ax^2+Bx+C=0.$$

15. Вывести формулы (5), (6), (7), (8), (9), (16), а также (5′), (6′), (7′), (8′), (9′), (17) из § 10.

16. Какому способу решения уравнения $x^3-(a^3+r)=0$ соответствует формула

$$\sqrt[3]{a^3+r} = a + \frac{r}{3a^2+3a+1},$$

которую предложил в 1202 г. Леонардо Пизанский для приближённого извлечения корней третьей степени из чисел?

17. Какому способу решения уравнений соответствует известный ещё вавилонским математикам способ извлечения корня квадратного из чисел

$$\sqrt{a^2+b^2} \approx a + \frac{b}{2a} \; ? (b<a),$$

а способ, предложенный арабским математиком ал-Кархи (XI в.):

$$\sqrt{a^2+r} = a + \frac{r}{2a+1} \; ?$$

18. Применить способ Ньютона к улучшению значения мнимого корня уравнения $x^2+x+1=0$, положив $x_1\approx -0{,}5+0{,}85\,i\approx a$. Применить для той же цели обобщённый способ Ньютона — формула (11) из § 8. Сравнить результаты с $x_1=-\frac{1}{2}+i\frac{\sqrt{3}}{2}$.

19. В каком виде следует взять уравнение для сходимости способа итерации

$$x=\operatorname{tg} x \quad \text{или} \quad \operatorname{arctg} x = x\,?$$

Вычислить наименьший положительный корень с точностью до 0,01, найдя табличным способом исходное приближение.

ГЛАВА III

РЕШЕНИЕ СИСТЕМ УРАВНЕНИЙ

§ 11. Способ Ньютона

Переходя к вопросу о решении систем уравнений, рассмотрим, прежде всего, способ Ньютона, одинаково пригодный как для алгебраических, так и для трансцендентных уравнений.

Возьмём для простоты систему из двух уравнений с двумя неизвестными:

$$\left.\begin{aligned} \varphi(x,\ y) &= 0, \\ \psi(x,\ y) &= 0. \end{aligned}\right\} \tag{1}$$

Пусть для интересующей нас пары действительных корней (x_1, y_1) найдены приближённые значения a и b; обозначим погрешности этих приближений через α и β:

$$\left.\begin{aligned} x_1 &= a + \alpha \approx a, \\ y_1 &= b + \beta \approx b. \end{aligned}\right\} \tag{2}$$

Разлагая функции $\varphi(x, y)$ и $\psi(x, y)$ в ряд Тейлора по степеням $x - a$ и $y - b$ и заменяя сразу в полученных разложениях x и y числами x_1 и y_1, получим:

$$\left.\begin{aligned} \varphi(x_1,\ y_1) = 0 &= \varphi(a,\ b) + \varphi'_x(a,\ b)(x_1 - a) + \\ &+ \varphi'_y(a,\ b)(y_1 - b) + \frac{1}{2!}[\varphi''_{xx}(a,\ b)(x_1 - a)^2 + \ldots] + \ldots, \\ \psi(x_1,\ y_1) = 0 &= \psi(a,\ b) + \psi'_x(a,\ b)(x_1 - a) + \\ &+ \psi'_y(a,\ b)(y_1 - b) + \frac{1}{2!}[\psi''_{xx}(a,\ b)(x_1 - a)^2 + \ldots] + \ldots \end{aligned}\right\} \tag{3}$$

Из этих уравнений надо найти $x_1 - a$ и $y_1 - b$ (т. е. α и β).

В способе Ньютона, пренебрегая в уравнениях (3) членами второй и более высоких степеней относительно $x_1 - a$ и $y_1 - b$, вместо системы (3) решают систему

$$\left.\begin{aligned} \varphi(a,\ b) + \varphi'_x(a,\ b)\alpha + \varphi'_y(a,\ b)\beta &= 0, \\ \psi(a,\ b) + \psi'_x(a,\ b)\alpha + \psi'_x(a,\ b)\beta &= 0, \end{aligned}\right\} \tag{4}$$

что, естественно, даёт приближённые значения $\bar{\alpha}$ и $\bar{\beta}$ искомых поправок α и β:

$$\bar{\alpha}=-\frac{\begin{vmatrix}\varphi(a,b) & \varphi'_y(a,b)\\ \psi(a,b) & \psi'_y(a,b)\end{vmatrix}}{\begin{vmatrix}\varphi'_x(a,b) & \varphi'_y(a,b)\\ \psi'_x(a,b) & \psi'_y(a,b)\end{vmatrix}},\quad \bar{\beta}=-\frac{\begin{vmatrix}\varphi'_x(a,b) & \varphi(a,b)\\ \psi'_x(a,b) & \psi(a,b)\end{vmatrix}}{\begin{vmatrix}\varphi'_x(a,b) & \varphi'_y(a,b)\\ \psi'_x(a,b) & \psi'_y(a,b)\end{vmatrix}}. \tag{5}$$

Прибавив $\bar{\alpha}$ и $\bar{\beta}$ к приближённым значениям a и b, получим:

$$x_1\approx a_1=a+\bar{\alpha},$$
$$y_1\approx b_1=b+\bar{\beta}.$$

Если погрешности этих приближённых равенств обозначим через α_1 и β_1, то для их определения надо будет решить уравнения, аналогичные (4):

$$\left.\begin{aligned}&\varphi(a_1,b_1)+\varphi'_x(a_1,b_1)\alpha_1+\varphi'_y(a_1,b_1)\beta_1=0,\\ &\psi(a_1,b_1)+\psi'_x(a_1,b_1)\alpha_1+\psi'_y(a_1,b_1)\beta_1=0,\end{aligned}\right\} \tag{6}$$

что даст нам $\bar{\alpha}_1$ и $\bar{\beta}_1$, а следовательно, и $a_2=a_1+\bar{\alpha}_1$, и $b_2=b_1+\bar{\beta}_1$, и т. д.

По мере приближения к искомым корням свободные члены $\varphi(a,b)$, $\psi(a,b)$, $\varphi(a_1,b_1)$, $\psi(a_1,b_1)$ и т. д., а следовательно, вообще говоря, и $\bar{\alpha}_k$, $\bar{\beta}_k$ будут неограниченно убывать по абсолютной величине, и вычисления можно будет прекратить, когда получаемые при решении уравнений типа (4) и (6) «поправки» будут достаточно малы (в пределах сохраняемых при вычислении знаков).

Действительно, если, например, $\bar{\alpha}_k$ и $\bar{\beta}_k$ настолько малы, что

$$a_{k+1}=a_k+\bar{\alpha}_k\approx a_k,$$
$$b_{k+1}=b_k+\bar{\beta}_k\approx b_k$$

(в пределах сохраняемых знаков), то уравнения для определения поправок $\bar{\alpha}_{k+1}$ и $\bar{\beta}_{k+1}$ ничем не будут отличаться от уравнений для α_k и β_k, следовательно, и новые поправки будут исчезающе малыми.

Пример. Рассмотрим систему

$$\left.\begin{aligned}&x^2+y-5=0,\\ &y^2+x-7=0.\end{aligned}\right\} \tag{7}$$

Если построим кривые, соответствующие этим уравнениям, то координаты точек пересечения полученных кривых будут удовлетворять каждому из этих уравнений, т. е. будут корнями системы (7).

Из рис. 6 видно, что система (7) имеет четыре пары корней. Если вычертить кривые на миллиметровке, то легко получить:

$$x_1 \approx -1{,}4 = a,$$
$$y_1 \approx +2{,}9 = b.$$

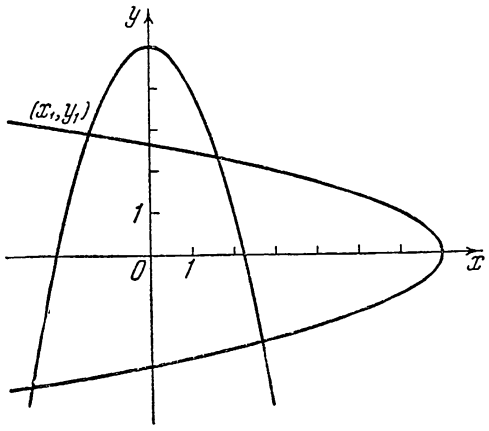

Рис. 6.

Так как

$$\left.\begin{aligned} &\varphi(x, y) = x^2 + y - 5; \\ &\psi(x, y) = y^2 + x - 7, \\ &\frac{\partial\varphi}{\partial x} = 2x, \quad \frac{\partial\varphi}{\partial y} = 1; \\ &\frac{\partial\psi}{\partial x} = 1, \quad \frac{\partial\psi}{\partial y} = 2y, \end{aligned}\right\} \quad (8)$$

то, вычислив значения этих функций при $x = a$ и при $y = b$, легко получить уравнения

$$\left.\begin{aligned} -0{,}14 - 2{,}8\,\alpha + \beta = 0, \\ 0{,}01 + \alpha + 5{,}8\,\beta = 0, \end{aligned}\right\} \quad (4')$$

откуда $\bar{\alpha} \approx -0{,}051$, $\bar{\beta} \approx 0{,}007$, следовательно,

$$x_1 \approx a_1 = a + \bar{\alpha} = -1{,}451, \quad y_1 \approx b_1 = b + \bar{\beta} = 2{,}907.$$

Вычислив значения функций (8) при $x = a_1$ и $y = b_1$, получим:

$$\left.\begin{aligned} 0{,}0124 - 2{,}902\,\alpha_1 + \beta_1 = 0, \\ -0{,}0004 + \alpha_1 + 5{,}814\,\beta_1 = 0, \end{aligned}\right\} \quad (6')$$

откуда $\bar{\alpha}_1 = +0{,}0041$, $\bar{\beta}_1 = -0{,}0006$, следовательно,

$$x_1 \approx a_2 = a_1 + \bar{\alpha}_1 = -1{,}4469 \approx -1{,}447,$$
$$y_1 \approx b_2 = b_1 + \bar{\beta}_1 = 2{,}9064 \approx 2{,}906.$$

Если бы мы вычислили поправки $\bar{\alpha}_2$ и $\bar{\beta}_2$, то они не изменили бы третьи десятичные знаки, т. е.

$$a_3 = -1{,}447 = a_2,$$

$$b_3 = 2{,}906 = b_2;$$

следовательно, эти числа можно принять за приближённые значения искомых корней.

Отметим, что при подобного рода вычислениях удобно пользоваться арифмометром.

Конечно, в том случае, когда из данной системы уравнений (1) просто исключается одно из неизвестных, дело можно свести к решению одного уравнения с одним неизвестным.

Например, для системы алгебраических уравнений такое исключение всегда возможно; при этом, если даны уравнения m-й и n-й степеней, то после исключения одного неизвестного согласно теореме Безу [1]) получается алгебраическое уравнение очень высокой степени, — вообще говоря, степени mn (правда, иногда старшие коэффициенты полученного уравнения могут оказаться равными нулю). Поэтому даже при отыскании действительных корней системы алгебраических уравнений практичнее применять изложенный выше способ Ньютона или способ итерации, рассмотрению которого будет посвящён следующий параграф.

Сущность способа Ньютона остаётся и для систем n уравнений с n неизвестными та же, что и для двух уравнений с двумя неизвестными: имея приближённые значения корней, поправки для них находят из системы линейных уравнений, которые получаются при отбрасывании членов второго и высших порядков в разложении в ряд Тейлора левых частей уравнений.

§ 12. Способ итерации

Пусть для интересующей нас пары корней (x_1, y_1) системы уравнений

$$\left.\begin{aligned} \varphi(x, y) &= 0, \\ \psi(x, y) &= 0 \end{aligned}\right\} \tag{1}$$

каким-нибудь образом (например, графическим или, более точно, способом Ньютона) найдены приближённые значения

$$\left.\begin{aligned} x_1 &\approx a_0, \\ y_1 &\approx b_0. \end{aligned}\right\} \tag{2}$$

Пусть, далее, система (1) заменена равносильной ей системой

$$\left.\begin{aligned} x &= \Phi(x, y), \\ y &= \Psi(x, y) \end{aligned}\right\} \tag{3}$$

и найдены пары чисел (a_1, b_1), (a_2, b_2), ..., (a_n, b_n), ... по формулам

$$\left.\begin{aligned} a_k &= \Phi(a_{k-1}, b_{k-1}), \\ b_k &= \Psi(a_{k-1}, b_{k-1}). \end{aligned}\right\} \tag{4}$$

---

[1]) О технике исключения неизвестного из системы двух алгебраических уравнений и о теореме Безу подробнее говорится в учебниках высшей алгебры.

Тогда имеет место

Теорема. *Если функции* $\Phi(x, y)$ *и* $\Psi(x, y)$ *таковы, что для всех точек прямоугольной области (с краями, параллельными осям координат), содержащей точки* $(x_1, y_1), (a, b), (a_1, b_1), \ldots, (a_n, b_n), \ldots$ *выполняются условия*

$$\left.\begin{aligned} \left|\frac{\partial\Phi}{\partial x}\right| + \left|\frac{\partial\Psi}{\partial x}\right| < m < 1, \\ \left|\frac{\partial\Phi}{\partial y}\right| + \left|\frac{\partial\Psi}{\partial y}\right| < m < 1, \end{aligned}\right\} \tag{5}$$

то при возрастании k *числа* $|x_1 - a_k|$ *и* $|y_1 - b_k|$ *могут быть сделаны как угодно малыми.*

Другими словами, при выполнении условий (5) последовательное вычисление по формулам (4) пар чисел (a_1, b_1), (a_2, b_2) и т. д. может дать значение искомых корней с любой степенью точности.

Доказательство. Для корней (x_1, y_1) системы (3) мы имеем точные соотношения

$$\left.\begin{aligned} x_1 = \Phi(x_1, y_1), \\ y_1 = \Psi(x_1, y_1). \end{aligned}\right\} \tag{6}$$

Вычитая почленно из первого равенства (6) первое равенство (4), а из второго — второе и применяя формулу Тейлора для функций от двух переменных, получим:

$$\begin{aligned} x_1 - a_k &= \Phi(x_1, y_1) - \Phi(a_{k-1}, b_{k-1}) = \\ &= (x_1 - a_{k-1})\bar{\Phi}'_x + (y_1 - b_{k-1})\bar{\Phi}'_y, \end{aligned} \tag{7}$$

$$\begin{aligned} y_1 - b_k &= \Psi(x_1, y_1) - \Psi(a_{k-1}\ b_{k-1}) = \\ &= (x_1 - a_{k-1})\bar{\Psi}'_x + (y_1 - b_{k-1})\Psi'_y, \end{aligned} \tag{8}$$

где через $\bar{\Phi}'_x$, $\bar{\Phi}'_y$, $\bar{\Psi}'_x$ и $\bar{\Psi}'_y$ обозначены, соответственно, значения функций $\frac{\partial\Phi}{\partial x}$, $\frac{\partial\Phi}{\partial y}$, $\frac{\partial\Psi}{\partial x}$ и $\frac{\partial\Psi}{\partial y}$ при некоторых значениях x и y, лежащих в промежутках (a_{k-1}, x_1) и (b_{k-1}, y_1).

Переход к абсолютным значениям в равенствах (7) и (8) даст нам:

$$|x_1 - a_k| \leqslant |x_1 - a_{k-1}| \cdot |\bar{\Phi}'_x| + |y_1 - b_{k-1}| \cdot |\bar{\Phi}'_y|, \tag{7'}$$

$$|y_1 - b_k| \leqslant |x_1 - a_{k-1}| \cdot |\bar{\Psi}'_x| + |y_1 - b_{k-1}| \cdot |\bar{\Psi}'_y|. \tag{8'}$$

Складывая почленно последние два неравенства, получим:

$$\begin{aligned} &|x_1 - a_k| + |y_1 - b_k| \leqslant \\ &\leqslant |x_1 - a_{k-1}|\{|\bar{\Phi}'_x| + |\bar{\Psi}'_y|\} + |y_1 - b_{k-1}|\{|\bar{\Phi}'_y| + |\bar{\Psi}'_y|\}, \end{aligned} \tag{9}$$

откуда в силу условий (5)

$$|x_1 - a_k| + |y_1 - b_k| < m\{|x_1 - a_{k-1}| + |y_1 - b_{k-1}|\}. \tag{10}$$

Если введём обозначение

$$|x_1 - a_s| + |y_1 - b_s| = \Delta_s, \tag{11}$$

то, придавая в (10) k значения 1, 2, 3, ..., $n-1$, n, получим:

$$\left.\begin{aligned} \Delta_1 &< m\Delta_0, \\ \Delta_2 &< m\Delta_1, \\ &\cdots\cdots \\ \Delta_{n-1} &< m\Delta_{n-2}, \\ \Delta_n &< m\Delta_{n-1}. \end{aligned}\right\} \tag{12}$$

Наконец, почленное перемножение соотношений (12) даёт после сокращения: $\Delta_n < m^n \Delta_0$ или

$$|x_1 - a_n| + |y_1 - b_n| < m^n \{|x_1 - a_0| + |y_1 - b_0|\}, \tag{13}$$

что и доказывает теорему, так как $\lim\limits_{n\to\infty} m^n = 0$ при $0 < m < 1$.

Как же, однако, преобразовать систему (1) к виду (3), чтобы при этом выполнялись условия (5)?

Иногда это удаётся сделать довольно просто, применяя к каждому из уравнений (1) тот или иной искусственный приём и пройдя, может быть, через ряд неудачных попыток.

Например, систему уравнений

$$\left.\begin{aligned} x^2 + y - 5 &= 0, \\ y^2 + x - 7 &= 0, \end{aligned}\right\} \tag{14}$$

если нас интересует пара корней: $x_1 \approx -1{,}4 = a_0$ и $y_1 \approx 2{,}9 = b_0$, можно преобразовать к виду

$$\left.\begin{aligned} x &= 7 - y^2, \\ y &= 5 - x^2 \end{aligned}\right\} \tag{14'}$$

или к виду

$$\left.\begin{aligned} x &= -\sqrt{5-y}, \\ y &= +\sqrt{7-x}, \end{aligned}\right\} \tag{14''}$$

но для системы (14') условия (5) не выполняются (так как, например, $\left|\frac{\partial \Phi}{\partial y}\right| = |-2y|$, что при $y = 2{,}9 = b_0$ даёт большое число 5,8). Для системы же (14'') при $x = a_0 = -1{,}4$ и $y = b_0 = 2{,}9$ имеем:

$$\frac{\partial \Phi}{\partial x} = 0, \quad \frac{\partial \Phi(a_0, b_0)}{\partial y} = \frac{1}{2\sqrt{5-2{,}9}}, \quad \text{т. е.} \quad \left|\frac{\partial \Phi(a_0, b_0)}{\partial y}\right| < 0{,}4,$$

$$\frac{\partial \Psi}{\partial y} = 0, \quad \frac{\partial \Psi(a_0, b_0)}{\partial x} = -\frac{1}{2\sqrt{7+1{,}4}}, \quad \text{т. е.} \quad \left|\frac{\partial \Psi(a_0, b_0)}{\partial x}\right| < 0{,}2.$$

Как видно будет из дальнейших вычислений, точки (a_1, b_1), (a_2, b_2) и т. д. и, наконец, (x_1, y_1) будут лежать близко около точки (a_0, b_0), а потому соответствующие частные производные функции $\Phi(x, y)$ и $\Psi(x, y)$ в прямоугольнике, о котором говорилось в теореме, не могут значительно превышать чисел 0,4 и 0,2.

Следовательно, условия (5) для системы (14″) соблюдаются, и по способу итерации можно с любой степенью точности вычислить искомые корни.

Пользуясь таблицами Барлоу (столбцом с заголовком $\sqrt{10n}$), найдём:

$a_1 = -\sqrt{5 - b_0} = -\sqrt{2{,}1} \approx -1{,}449$, $\quad a_3 = -\sqrt{2{,}093} \approx -1{,}447$,

$b_1 = \sqrt{7 - a_0} = -\sqrt{8{,}4} \approx 2{,}898$, $\quad b_3 = \sqrt{8{,}450} \approx 2{,}907$,

$a_2 = -\sqrt{2{,}102} \approx -1{,}450$, $\quad a_4 = -\sqrt{2{,}093} \approx -1{,}447$,

$b_2 = \sqrt{8{,}449} \approx 2{,}907$, $\quad b_4 = \sqrt{8{,}447} \approx 2{,}906$.

Следующий шаг нам даёт:

$$a_5 = \Phi(a_4, b_4) \approx -1{,}447 = a_4,$$
$$b_5 = \Psi(a_4, b_4) \approx 2{,}906 = b_4,$$

т. е. в пределах точности вычислений можно считать, что a_4 и b_4 удовлетворяют уравнениям (14″), т. е.

$$x_1 \approx -1{,}447, \quad y_1 \approx 2{,}906.$$

В общем случае, если при решении системы (3) получим

$$a_k = \Phi(a_{k-1}, b_{k-1}) \approx a_{k-1},$$
$$b_k = \Psi(a_{k-1}, b_{k-1}) \approx b_{k-1},$$

то

$$x_1 \approx a_{k-1} \quad \text{и} \quad y_1 \approx b_{k-1},$$

так как эти числа удовлетворяют (в пределах точности вычислений) системе уравнений (3). Но далеко не всегда удаётся так просто добиться выполнения условий (5) путём преобразования каждого из данных уравнений в отдельности. Однако, как легко убедиться, комбинируя данные уравнения друг с другом, почти всегда можно добиться выполнения условий (5) при достаточно малом значении m, что обеспечит быструю сходимость процесса.

При этом надо оговориться, что исходное приближение $x_1 \approx a_0$, $y_1 \approx b_0$ должно быть тем лучше, чем бóльшие (по абсолютной величине) значения имеют вторые производные функций $\varphi(x, y)$ и $\psi(x, y)$ в области, заключающей точки (a_0, b_0), (a_1, b_1) и т. д.

Действительно, легко видеть, что системе уравнений

$$\left.\begin{aligned} \alpha\varphi(x, y) + \beta\psi(x, y) &= 0, \\ \gamma\varphi(x, y) + \delta\psi(x, y) &= 0 \end{aligned}\right\} \qquad (15)$$

удовлетворяет любая пара корней (x_i, y_i) системы (1), каковы бы ни были постоянные α, β, γ, δ.

Если же эти числа удовлетворяют условию

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma \neq 0, \tag{16}$$

то левые части (15) одновременно могут обращаться в нуль лишь при условии, что будут обращаться в нуль $\varphi(x, y)$ и $\psi(x, y)$ [так как система n однородных линейных уравнений с n неизвестными при определителе, отличном от нуля, может иметь только «нулевое» решение [1])]. Следовательно, при условии (16) любое решение системы уравнений (15) будет в то же время решением системы (1), т. е. системы уравнений (15) и (1) будут эквивалентны.

Перепишем систему (15) в виде

$$\left.\begin{aligned} x &= x + \alpha\varphi(x, y) + \beta\psi(x, y) = \Phi(x, y), \\ y &= y + \gamma\varphi(x, y) + \delta\psi(x, y) = \Psi(x, y) \end{aligned}\right\} \tag{15'}$$

и подберём постоянные α, β, γ, δ так, чтобы производные правых частей равенств (15') при $x = a_0$ и $y = b_0$ были бы равны нулю, т. е. найдём α и β из уравнений

$$\left.\begin{aligned} \Phi'_x(a_0, b_0) &= 1 + \alpha\varphi'_x(a_0, b_0) + \beta\psi'_x(a_0, b_0) = 0, \\ \Phi'_y(a_0, b_0) &= \alpha\varphi'_y(a_0, b_0) + \beta\psi'_y(a_0, b_0) = 0, \end{aligned}\right\} \tag{17}$$

а γ и δ — из уравнений

$$\left.\begin{aligned} \Psi'_x(a_0, b_0) &= \gamma\varphi'_x(a_0, b_0) + \delta\psi'_x(a_0, b_0) = 0, \\ \Psi'_y(a_0, b_0) &= 1 + \gamma\varphi'_y(a_0, b_0) + \delta\psi'_y(a_0, b_0) = 0. \end{aligned}\right\} \tag{18}$$

Если $\varphi'_x(x, y)$, $\varphi'_y(x, y)$, $\psi'_x(x, y)$ и $\psi'_y(x, y)$ в окрестности точки (a_0, b_0) меняются не очень быстро, т. е. вторые производные $\varphi''_{xx}(a_0, b_0)$, $\varphi''_{xy}(a_0, b_0)$ и т. п. достаточно малы, а числа a_0, b_0 достаточно близки к искомым корням, то $\Phi'_x(x, y)$, $\Phi'_y(x, y)$, $\Psi'_x(x, y)$, $\Psi'_y(x, y)$ будут иметь в интересующей нас области значения, мало отличающиеся от нуля, т. е. условия (5) будут выполняться.

Возьмём для примера уже рассмотренную систему

$$\left.\begin{aligned} \varphi(x, y) &= x^2 + y - 5 = 0, \\ \psi(x, y) &= y^2 + x - 7 = 0. \end{aligned}\right\}$$

Так как

$$\varphi'_x = 2x, \quad \varphi'_y = 1, \quad x_1 \approx -1{,}4 = a_0,$$
$$\psi'_x = 1, \quad \psi'_y = 2y, \quad y_1 \approx 2{,}9 = b_0,$$

---

[1]) См., например, стр. 73, Л. Я. Окунев, Кольцо многочленов и поле рациональных функций.

то уравнения (17) примут вид

$$1-2{,}8\alpha+\beta=0,\quad \alpha+5{,}8\beta=0. \tag{17'}$$

Следовательно,

$$\alpha=\frac{\begin{vmatrix}-1 & 1\\ 0 & 5{,}8\end{vmatrix}}{\begin{vmatrix}-2{,}8 & 1\\ 1 & 5{,}8\end{vmatrix}}=\frac{-5{,}8}{-17{,}24}\approx 0{,}34,\quad \beta=\frac{\begin{vmatrix}-2{,}8 & -1\\ 1 & 0\end{vmatrix}}{-17{,}24}\approx -0{,}06.$$

Уравнения (18) дадут:

$$-2{,}8\gamma+\delta=0,\quad 1+\gamma+5{,}8\delta=0. \tag{18'}$$

Следовательно,

$$\gamma=\frac{\begin{vmatrix}0 & 1\\ -1 & 5{,}8\end{vmatrix}}{\begin{vmatrix}-2{,}8 & 1\\ 1 & 5{,}8\end{vmatrix}}=\frac{1}{-17{,}24}\approx -0{,}06,\quad \delta=\frac{\begin{vmatrix}-2{,}8 & 0\\ 1 & -1\end{vmatrix}}{-17{,}24}\approx -0{,}16,$$

и система (15') в данном случае будет иметь вид

$$\left.\begin{aligned} x&=x+0{,}34\,(x^2+y-5)-0{,}06\,(y^2+x-7)=\Phi(x,\ y),\\ y&=y-0{,}06\,(x^2+y-5)-0{,}16\,(y^2+x-7)=\Psi(x,\ y).\end{aligned}\right\} \tag{15''}$$

Подставляя в правые части этих уравнений $x=a_0=-1{,}4$ и $y=b_0=2{,}9$, получим:

$$a_1=\Phi(a_0,\ b_0)=-1{,}4+0{,}34\cdot(-0{,}14)-0{,}06\cdot 0{,}01=-1{,}4482,$$

$$b_1=\Psi(a_0,\ b_0)=2{,}9-0{,}06\cdot(-0{,}14)-0{,}16\cdot 0{,}01=2{,}9068,$$

т. е. в один шаг достигнута почти такая же точность, на которую при первом способе решения потребовалось два-три шага.

Отметим, что нет необходимости корни систем линейных уравнений (17) и (18) вычислять с большой точностью, так как будут ли $\Phi'_x(a_0,\ b_0)$, $\Psi'_x(a_0,\ b_0)$ и т. п. точно равны нулю или почти равны нулю, — всё равно значения этих производных в точках $(a_1,\ b_1)$, $(a_2,\ b_2)$ и т. д. будут несколько отличаться от нуля, и неточное решение уравнений (17) и (18) мало скажется на условиях (5).

Может даже случиться, что отклонение от нуля функций $\Phi'_x(x,\ y)$, $\Phi'_y(x,\ y)$ и т. п. в точке $(a_0,\ b_0)$ вызовет уменьшение значений этих функций в точках $(a_1,\ b_1)$, $(a_2,\ b_2)$ и т. д. и в конечном счёте отзовётся благоприятно на значении m в условиях (5), т. е. на скорости приближения к точным корням системы уравнений (1). В силу этого имеет смысл определять α, β, γ, δ графическим путём; подробнее об этом будет сказано в главе IV.

Отметим, кстати, что способ итерации обладает замечательным свойством: если в процессе вычислений были допущены ошибки, например, при вычислении a_3, мы получили немного отличающееся от него число $\overline{a_3}$, то ошибка эта, если она невелика, может в худ-

шем случае замедлить приближение к искомым корням уравнения (если она удаляет нас от x_1); однако сделанная ошибка может случайно и приблизить нас к x_1, что сократит последующие вычисления. Опытный вычислитель, подметив закономерность в последовательных приближениях, может даже намеренно «ошибаться», подправляя «на-глаз» вычисляемые по формулам (4) значения (a_k, b_k).

Конечно, не следует думать, что вычисления в способе итерации можно проводить небрежно, надеясь на возможность благоприятного влияния ошибок, тем более, что ошибка может нас даже вывести за пределы той области, в которой выполняются условия (5). Естественно, что заключительный этап вычислений, при котором устанавливается, что числа (a_n, b_n) совпадают в пределах точности вычислений с (a_{n-1}, b_{n-1}), должен быть проведён со всей тщательностью, и ошибки на этом этапе абсолютно недопустимы.

Мы изложили способ итерации применительно к системе из двух уравнений с двумя неизвестными; однако ничего по существу не изменится при решении n уравнений с n неизвестными. Например, если три уравнения написаны в виде

$$\left.\begin{aligned} x &= F(x, y, z), \\ y &= \Phi(x, y, z), \\ z &= \Psi(x, y, z) \end{aligned}\right\} \tag{19}$$

и в прямоугольном параллелепипеде, заключающем точки (a_0, b_0, c_0), (a_1, b_1, c_1), ..., (x_1, y_1, z_1), соблюдаются условия

$$\left.\begin{aligned} |F'_x| + |\Phi'_x| + |\Psi'_x| &< m < 1, \\ |F'_y| + |\Phi'_y| + |\Psi'_y| &< m < 1, \\ |F'_z| + |\Phi'_z| + |\Psi'_z| &< m < 1, \end{aligned}\right\} \tag{20}$$

то по формулам

$$\left.\begin{aligned} a_k &= F(a_{k-1}, b_{k-1}, c_{k-1}), \\ b_k &= \Phi(a_{k-1}, b_{k-1}, c_{k-1}), \\ c_k &= \Psi(a_{k-1}, b_{k-1}, c_{k-1}) \end{aligned}\right\} \tag{21}$$

будем получать всё лучшие и лучшие приближения к корням (x_1, y_1, z_1) системы (19), если исходное приближение было:

$$x_1 \approx a_0, \quad y_1 \approx b_0, \quad z_1 \approx c_0.$$

Конечно, наиболее сложным является получение достаточно точного исходного приближения, которое обеспечило бы преобразование системы уравнений, данной в произвольном виде, к виду (19) с соблюдением условий (20); однако, как увидим в главе IV, графический способ позволяет иногда и эту задачу решить довольно просто (для уравнений с двумя или тремя неизвестными).

§ 13. Замечания о вычислении мнимых корней алгебраических уравнений

Вычисление мнимых корней уравнения

$$f(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n=0, \tag{1}$$

можно свести к нахождению действительных корней системы уравнений с двумя неизвестными.

Действительно, положим

$$x=u+vi; \tag{2}$$

тогда

$$f(x)=f(u+vi)=a_0(u+vi)^n+a_1(u+vi)^{n-1}+\ldots+a_n.$$

Если в правой части раскрыть скобки и собрать отдельно все действительные и все мнимые слагаемые, то получим:

$$f(x)=\varphi(u,\ v)+i\psi(u,\ v)=0. \tag{3}$$

Следовательно, u и v должны удовлетворять системе уравнений

$$\left.\begin{aligned}\varphi(u,\ v)&=0,\\ \psi(u,\ v)&=0.\end{aligned}\right\} \tag{4}$$

Приём этот пригоден и для уравнений с комплексными коэффициентами. Например, уравнение

$$x^4+(4+i)x+(-2+26i)=0 \tag{5}$$

после подстановки $x=u+vi$ принимает вид

$$(u^4+4u^3vi-6u^2v^2-4uv^3i+v^4)+(4+i)(u+vi)-2+26i=0$$

и распадается на два уравнения:

$$\left.\begin{aligned}u^4-6u^2v^2+v^4+4u-v-2&=0,\\ 4u^3v-4uv^3+u+4v+26&=0.\end{aligned}\right\} \tag{6}$$

Легко, например, проверить, что одно из решений системы (6) $u_1=2$, $v_1=-1$ даёт мнимый корень $x_1=2-i$ уравнения (5).

Конечно, даже при небольших значениях n получается весьма сложная система уравнений вида (4), так что приём этот интересен главным образом в методическом отношении. Правда, если приближённое значение мнимого корня $x_1\approx a+bi$ получено заранее каким-нибудь путём [например, по способу Лобачевского, если уравнение (1) имеет действительные коэффициенты], то решение системы (4) упрощается, так как, считая $u_1\approx a$ и $v_1\approx b$ за исходное приближение, остаётся только улучшить его, что можно сделать хотя бы по способу Ньютона (§ 11).

Отметим ещё, что, зная приближённое значение мнимого корня уравнения (1) $x_1 \approx \alpha = a + bi$, можно улучшить его и по формуле

$$\alpha_1 = \alpha - \frac{f(\alpha)}{f'(\alpha)},$$

соответствующей способу Ньютона для уравнения с одним неизвестным [1]).

Задачи к главе III

20. Решить по способу Ньютона и по способу итерации следующие системы, вычисляя указанные корни с точностью до 0,001:

а) $2x^3 - y^2 - 1 = 0$, $xy^3 - y - 4 = 0$, $x_1 \approx 1{,}2$, $y_1 \approx 1{,}7$;

б) $x + 3\lg x - y^2 = 0$, $2x^2 - xy - 5x + 1 = 0$, $x_1 \approx 1{,}4$, $y_1 \approx -1{,}5$, $x_2 \approx 3{,}4$, $y_2 \approx 2{,}2$.

21. Вычислить с точностью до 0,001 четыре пары корней системы

$$x^2 + y = 31, \quad y^2 + x = 41,$$

найдя исходные приближения графически.

22. Вычислить с точностью до 0,001 корни x_1 и y_1 системы

$$x = \operatorname{tg} y, \quad y = \operatorname{tg} x,$$

приняв за исходное приближение

$$x_1 \approx 1{,}33, \quad y_1 \approx 4{,}07.$$

23. Показать, что действительная часть мнимого корня $x = u + vi$ уравнения третьей степени

$$f(x) = a_0x^3 + a_1x^2 + a_2x + a_3 = 0$$

должна удовлетворять уравнению

$$f(u)f'''(u) - 3f'(u)f''(u) = 0.$$

24. Вычислить по способу Ньютона и по способу итерации с точностью до 0,01 корни системы

$$x - y + 3z - 7 = 0, \quad x^2 + y^2 - z^2 = 0, \quad x - y^2 + 2z - 2{,}5 = 0,$$

приняв

$$x_1 \approx 1, \quad y_1 \approx -1{,}3, \quad z_1 \approx 1{,}6.$$

---

[1]) См., например, А. Н. Крылов, Лекции о приближённых вычислениях, изд. 5-е, Гостехиздат, 1950.

ГЛАВА IV

ГРАФИЧЕСКИЕ МЕТОДЫ

§ 14. Уравнения с одним неизвестным

Графические методы решения уравнений дают, как правило, грубо приближённые значения корней; но если требуется бóльшая точность, то, взяв полученные значения за исходные приближения, можно их уточнить, например, по способу итерации или по способу Ньютона.

Если уравнение дано в виде

$$f(x)=0, \tag{1}$$

то для определения действительных корней его достаточно построить график функции $f(x)$, т. е. кривую $y=f(x)$, и найти абсциссы $x_1, x_2, \ldots$ точек пересечения этой кривой с осью Ox; эти абсциссы и будут корнями уравнения (1), так как для них $f(x_i)=y_i=0$.

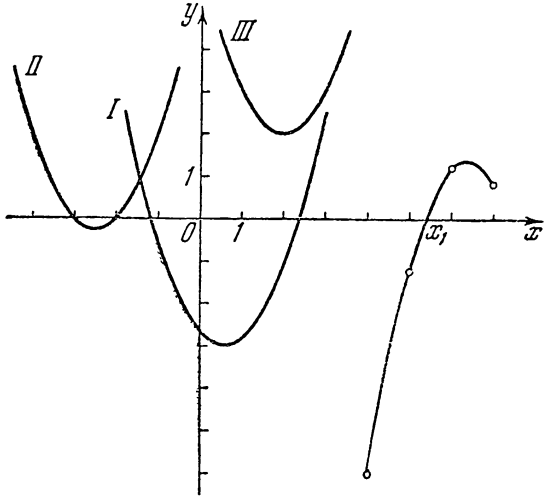

Рис. 7.

В некоторых случаях график $f(x)$ строится легко; например, для решения уравнения $ax-b=0$ (при a и b громоздких) достаточно построить по двум не очень близким точкам прямую $y=ax-b$.

Иногда бывает полезен прозрачный транспарант — начерченная на прозрачной бумаге специальная кривая. Например, для решения квадратных уравнений $x^2+px+q=0$ транспарант с кривой $y=x^2$ даёт сразу корни уравнения, если

вершину «прозрачной» параболы поместить в точку $\left(-\frac{p}{2}, q-\frac{p^2}{4}\right)$ чертежа, а ось её направить параллельно Oy.

Действительно,

$$f(x)=x^2+px+q=\left(x+\frac{p}{2}\right)^2+q-\frac{p^2}{4};$$

следовательно,

$$q-\frac{p^2}{4}=f\left(-\frac{p}{2}\right)$$

есть минимум функции $f(x)$.

Уравнение «прозрачной» параболы, расположенной указанным образом, по отношению к осям на чертеже будет:

$$y-\left(q-\frac{p^2}{4}\right)=\left(x+\frac{p}{2}\right)^2,$$

т. е.

$$y=x^2+px+q.$$

На рис. 7 показано решение уравнений

$$\text{I. } x^2-1{,}2x-2{,}6=0,$$

$$-\frac{p}{2}=0{,}6,\quad q-\frac{p^2}{4}\approx-3,\quad x_1\approx 2{,}4,\quad x_2\approx-1{,}2;$$

$$\text{II. } x^2+5x+6=0,$$

$$-\frac{p}{2}=-\frac{5}{2},\quad q-\frac{p^2}{4}=-\frac{1}{4},\quad x_1=-3,\quad x_2=-2;$$

$$\text{III. } x^2-4x+6=0,$$

$$-\frac{p}{2}=2,\quad q-\frac{p^2}{4}=2.$$

Так как «прозрачная» парабола уравнения III не пересекает оси Ox, то уравнение это не имеет действительных корней.

Для более сложных уравнений вблизи предполагаемого корня следует найти несколько точек кривой $y=f(x)$ (ниже и выше оси Ox) и провести через них от руки плавную кривую. Казалось бы, уже из самой таблицы можно найти, хотя бы по способу линейного интерполирования, приближённое значение корня, но иногда это грозит большой погрешностью, которую можно устранить построением графика; например, если таблица имеет вид

x	3	4	5	6	7
$f(x)$	−12	−6	−1,2	1,2	0,8

то способ линейного интерполирования подсказывал бы нам, что $x_1\approx 5{,}5$; график же, где мы приближённо осуществляем параболическое интерполирование, даёт нам $x_1\approx 5{,}4$ (см. рис. 7).

Иногда бывает полезно представить уравнение (1) в виде

$$F_1(x)=F_2(x) \qquad (2)$$

и, построив на одном чертеже (рис. 8) графики функций $F_1(x)$ и $F_2(x)$, найти абсциссы x_1, x_2 и т. д. точек пересечения полученных линий; найденные x_1, x_2 и т. д. и будут корнями уравнения (2) или (1), так как при любом из них

$$F_1(x_k) = F_2(x_k).$$

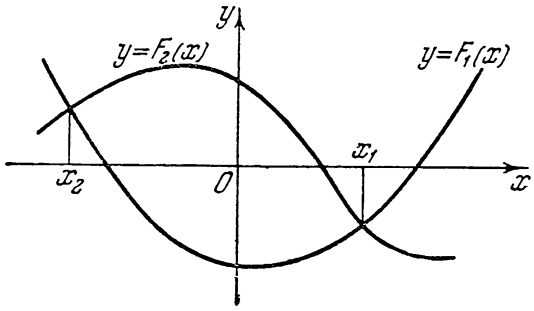

Рис. 8.

Приём этот выгодно применять, когда приходится решать ряд однотипных уравнений, для которых функция $F_1(x)$ — одна и та же, а функции $F_2(x)$ хотя и отличаются друг от друга, но их графики строятся просто. Тогда, построив тщательно один раз кривую $y = F_1(x)$, наносят на тот же чертёж кривые $y = F_2(x)$, соответствующие данным уравнениям.

Например, для квадратных уравнений вида

$$x^2 + ax + b = 0$$

имеем:

$$x^2 = -ax - b;$$

если на чертеже имеется тщательно построенная парабола $y = x^2$, то остаётся построить при данных a и b по двум точкам прямую $y = -ax - b$ и сразу «прочитать» на чертеже приближённые значения корней.

На рис. 9 показано решение уравнений

I. $x^2 - 1{,}5x - 3{,}5 = 0$, $x_1 \approx 2{,}8$, $x_2 \approx -1{,}3$;

II. $x^2 + 0{,}8x + 1{,}4 = 0$,

действительных корней нет, так как прямая $y = -0{,}8x - 1{,}4$ не пересекается с параболой $y = x^2$;

III. $z^2 + 0{,}02z - 0{,}1 = 0$, подстановка $z = \frac{x}{10}$ даёт $x^2 + 0{,}2x - 10 = 0$, на чертеже находим:

$$x_1 \approx 3{,}1, \quad x_2 \approx -3{,}3,$$

следовательно,

$$z_1 \approx 0{,}31, \quad z_2 \approx -0{,}33.$$

Тот же путь пригоден для уравнений вида

$$x^3 + px + q = 0$$

[здесь $F_1(x) = x^3$].

Для уравнений вида

$$z^3 + az^2 + bz + c = 0$$

удобна предварительная подстановка $z=-ax$, которая даст

$$x^3=x^2-\frac{b}{a^2}x+\frac{c}{a^3},$$

следовательно, имея на чертеже кубическую параболу $y=x^3$, мы, наложив соответствующим образом на чертёж «прозрачную» параболу $y=x^2$, сразу сможем «прочитать» значения x_1, x_2, x_3, а следовательно, и

$$z_k=-ax_k$$
$$(k=1,\ 2,\ 3).$$

Для уравнений типа

$$\sin kz-az-b=0$$

подстановка $z=\frac{x}{k}$ даёт:

$$\sin x=\frac{a}{k}x+b;$$

нанося на чертёж с заранее построенной синусоидой $y=\sin x$ прямые $y=\frac{a}{k}x+b$, сразу найдём x_1, x_2, ..., а следовательно, и z_1, z_2, ...

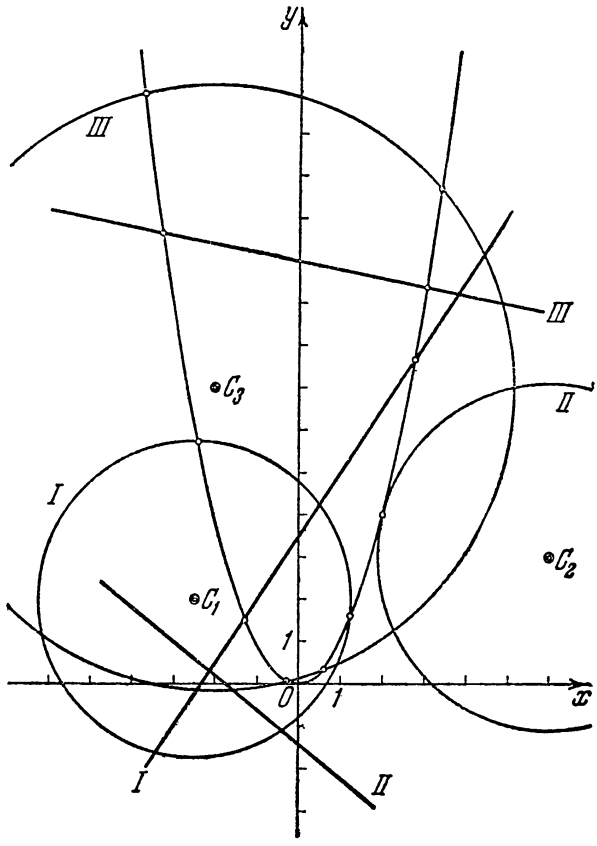

Рис. 9.

Замена уравнения (1) уравнением (2) равносильна решению системы уравнений

$$\left.\begin{aligned} y&=F_1(x),\\ y&=F_2(x),\end{aligned}\right\} \qquad (3)$$

при решении которой нас интересуют лишь значения корня x (соответствующие им значения y нам не нужны). Иногда удаётся заменить уравнение (1) простой системой более общего вида

$$\left.\begin{aligned} \varphi(x,\ y)&=0,\\ \psi(x,\ y)&=0,\end{aligned}\right\} \qquad (4)$$

которая при исключении неизвестного y приводит нас к уравнению (1). Например, для уравнения

$$x^4+ax^2+bx+c=0 \qquad (5)$$

можно взять систему

$$\left.\begin{aligned} x^2 - y &= 0, \\ (x-\alpha)^2 + (y-\beta)^2 - r^2 &= 0. \end{aligned}\right\} \tag{6}$$

Исключая из (6) y, получим:

$$(x-\alpha)^2 + (x^2-\beta)^2 - r^2 = \\ = x^4 + x^2(1-2\beta) - 2\alpha x + \alpha^2 + \beta^2 - r^2 = 0. \tag{7}$$

Если взять

$$\left.\begin{aligned} \alpha &= -\frac{b}{2}, \\ \beta &= \frac{1-a}{2}, \\ r &= \sqrt{\alpha^2 + \beta^2 - c}, \end{aligned}\right\} \tag{8}$$

то уравнение (7) совпадает с (5), и система (6) даст корни (x_1, y_1), (x_2, y_2) и т. д., где $x_1, x_2, \ldots$ будут корнями уравнения (5).

Так как графиками уравнений (6) будут парабола $y = x^2$ и окружность с центром $C(\alpha, \beta)$ и радиусом $r = \sqrt{\alpha^2 + \beta^2 - c}$, то получаем изящный способ решения уравнения четвёртой степени, понятный из рис. 9, где дано решение уравнений:

I. $x^4 - 3x^2 + 5x - 4 = 0$, $\alpha = -2{,}5$, $\beta = 2$, $r = \sqrt{6{,}25 + 4 + 4} \approx \approx 3{,}73$, $x_1 \approx -2{,}4$, $x_2 \approx 1{,}2$, $x_{3,4}$ — мнимые.

II. $x^4 - 5x^2 - 12x + 28 = 0$, $\alpha = 6$, $\beta = 3$, $r \approx 4{,}1$, $x_{1,2} = 2$ (кратный корень), $x_{3,4}$ — мнимые.

III. $x^4 - 13x^2 + 4x + 2 = 0$, $\alpha = -2$, $\beta = 7$, $r \approx 7{,}15$, $x_1 \approx -3{,}7$, $x_2 \approx -0{,}3$, $x_3 \approx 0{,}6$, $x_4 \approx 3{,}4$.

Решение алгебраических уравнений (способ Лилля). Удобным приёмом для быстрого решения алгебраических уравнений с действительными коэффициентами является способ, предложенный в 1876 г. французским инженером Лиллем. Сущность его поясним на примере уравнения четвёртой степени, что нисколько не уменьшит общности наших рассуждений.

Итак, пусть дано уравнение

$$a_0 x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0. \tag{9}$$

Левую часть этого уравнения [обозначим её для краткости через $F(x)$] представим в виде

$$F(x) = \{[(a_0 x + a_1)x + a_2]x + a_3\}x + a_4. \tag{10}$$

Нетрудно при конкретном значении $x=x_0$ последовательно построить отрезки, равные (при выбранной единице масштаба) выражениям

$$a_0x_0+a_1,\quad (a_0x_0+a_1)x_0+a_2,$$
$$[(a_0x_0+a_1)x_0+a_2]x_0+a_3,\quad F(x_0).$$

Для этого надо построить так называемую схему многочлена $F(x)$; делается это так.

Выбрав произвольную начальную точку O и единицу масштаба, последовательно откладывают отрезки:

$OA_0=a_0$	—	он откладывается	вверх	при	$a_0>0$	и вниз	при	$a_0<0$,
$A_0A_1=a_1$	—	»	вправо	»	$a_1>0$	и влево	»	$a_1<0$,
$A_1A_2=a_2$	—	»	вниз	»	$a_2>0$	и вверх	»	$a_2<0$,
$A_2A_3=a_3$	—	»	влево	»	$a_3>0$	и вправо	»	$a_3<0$,
$A_3A_4=a_4$	—	»	вверх	»	$a_4>0$	и вниз	»	$a_4<0$.

Ломаная $OA_0A_1A_2A_3A_4$ и будет схемой многочлена $F(x)$.

Для уравнений более высоких степеней отрезки для a_5, a_6, a_7, a_8,... откладываются, соответственно, так же, как для a_1, a_2, a_3, a_4,...

Если какой-нибудь коэффициент равен нулю, например $a_k=0$, то длина соответствующего отрезка $A_{k-1}A_k$ также будет равна нулю, т. е. точки A_{k-1} и A_k совпадут.

На рис. 10 приведена схема для уравнения

$$F(x)=x^4+3x^3-x^2-4x+4=0. \quad (11)$$

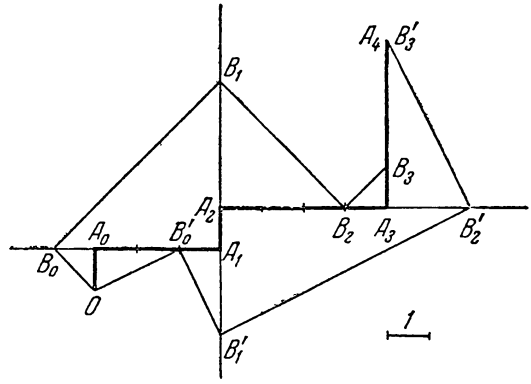

Рис. 10.

Определив затем при заданном x_0 такой угол φ, лежащий в промежутке $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, для которого $\operatorname{tg}\varphi=x_0$, строят ломаную $OB_0B_1B_2B_3$, проводя:

OB_0 под углом φ к отрезку OA_0 до пересечения с A_0A_1,

затем проводя:

$B_0B_1\perp OB_0$ до пересечения с A_1A_2,
$B_1B_2\perp B_0B_1$ » » » A_2A_3,
$B_2B_3\perp B_1B_2$ » » » A_3A_4

(при $\varphi > 0$ угол откладывается против вращения часовой стрелки, а при $\varphi < 0$ — по направлению её вращения).

Тогда из рис. 10 видно, что

$$\left.\begin{aligned}
&\overline{B_0A_0} = \overline{OA_0}\,\mathrm{tg}\,\varphi = a_0x_0,\\
&\overline{B_0A_1} = \overline{B_0A_0} + \overline{A_0A_1} = a_0x_0 + a_1,\\
&\overline{B_1A_1} = \overline{B_0A_1}\,\mathrm{tg}\,\varphi = (a_0x_0 + a_1)\,x_0,\\
&\overline{B_1A_2} = \overline{B_1A_1} + \overline{A_1A_2} = (a_0x_0 + a_1)\,x_0 + a_2,\\
&\overline{B_2A_2} = \overline{B_1A_2}\,\mathrm{tg}\,\varphi = [(a_0x_0 + a_1)\,x_0 + a_2]\,x_0,\\
&\overline{B_2A_3} = \overline{B_2A_2} + \overline{A_2A_3} = [(a_0x_0 + a_1)\,x_0 + a_2]\,x_0 + a_3,\\
&\overline{B_3A_3} = \overline{B_2A_3}\,\mathrm{tg}\,\varphi = \{[(a_0x_0 + a_1)\,x_0 + a_2]\,x_0 + a_3\}\,x_0,\\
&\overline{B_3A_4} = \overline{B_3A_3} + \overline{A_3A_4} = F(x_0).
\end{aligned}\right\} \quad (12)$$

На рис. 10 показано вычисление $F(1)$ с помощью ломаной $OB_0B_1B_2B_3$ (для неё $\mathrm{tg}\,\varphi = 1$) и $F(-2)$ с помощью ломаной $OB_0'B_1'B_2'B_3'$ ($\mathrm{tg}\,\varphi_1 = -2$). Мы имеем:

$$\overline{B_3A_4} = +3 = F(1),$$

что совпадает с результатами подстановки $x = 1$ в левую часть уравнения (11); точка же B_3' совпала с A_4, следовательно, $F(-2) = \overline{B_3'A_4} = 0$, т. е. $x = -2$ есть корень уравнения (11).

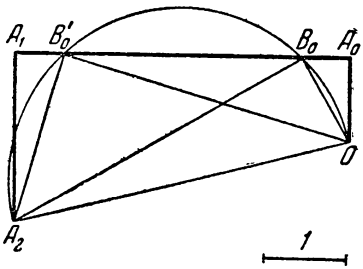

Рис. 11.

Следует обратить внимание на то, что в равенствах типа $\overline{B_kA_{k+1}} = \overline{B_kA_k} + \overline{A_kA_{k+1}}$ [см. (12)] мы имеем дело со сложением направленных отрезков, а в равенствах типа $\overline{B_kA_k} = \overline{B_{k-1}A_k}\,\mathrm{tg}\,\varphi$ знак произведения находится в строгом соответствии со знаками множителей, в чём читатель без труда убедится по рис. 10.

Значение угла φ, дающее корень уравнения, можно найти в общем случае путём нескольких проб, хотя удобнее всего воспользоваться для этой цели прозрачной бумагой, разлинованной в мелкую клетку: наложив такой транспарант на схему уравнения $OA_0A_1A_2A_3A_4$ под некоторым углом к основным линиям, легко проследить глазом, куда упадут точки B_0, B_1, B_2, B_3; вращая же транспарант, легко найти такие положения, при которых точка B_3 совпадёт с A_4.

Особенно просто находятся подходящие направления для OB_0 при решении квадратных уравнений, так как в этом случае достаточно найти на среднем звене A_0A_1 схемы $OA_0A_1A_2$ (или на его продолжении) точку B_0, из которой отрезок OA_2 был бы виден под

прямым углом; следовательно, строя окружность на диаметре OA_2, получим точки B_0 и B_0', дающие корни уравнения.

На рис. 11 показано решение уравнения

$$x^2 - 4x + 2 = 0.$$

При решении уравнений третьей степени невозможно в общем случае, пользуясь циркулем и линейкой, построить ломаную $OB_0B_1B_2$, дающую корень уравнения. Однако, как видно из рис. 12, пользуясь двумя «прямыми углами», можно, меняя их положение, добиться совпадения точки B_2 с точкой A_3. В этом смысле и говорят, что задачи на построение третьей степени, в которых искомый отрезок есть корень некоторого уравнения третьей степени (например, знаменитые задачи о трисекции угла и об удвоении куба), могут быть решены с помощью двух прямых углов.

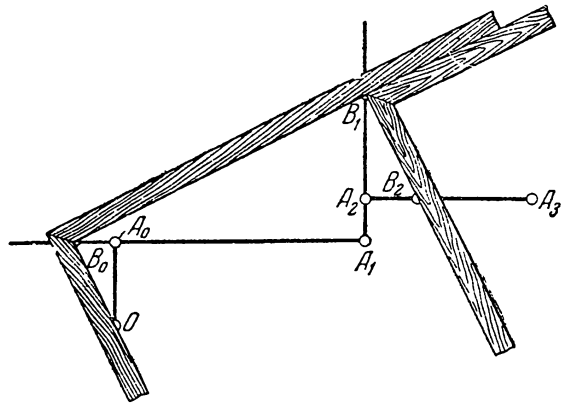

Рис. 12.

§ 15. Решение уравнений с помощью номограмм

Одной из разновидностей графического метода решения уравнений является использование номограмм.

Номограммой называют специальный чертёж, построенный для уравнений определённого типа и позволяющий быстро находить действительные корни любого уравнения этого типа.

Так, например, можно построить номограммы для решения уравнений следующих типов:

$$x^2 + px + q = 0, \tag{1}$$

$$x^3 + px + q = 0, \tag{2}$$

$$x^n - a = 0, \tag{3}$$

$$x^3 + ax^2 + bx + c = 0,$$

и т. д.

Чем меньше параметров входит в уравнение, тем проще будет и номограмма, тем быстрее при известных значениях параметров, соответствующих данному конкретному уравнению, можно найти его корни.

Мы ограничимся рассмотрением лишь простейших номограмм для уравнений с двумя параметрами типа (1), (2), (3), отсылая за

подробностями к специальным учебникам номографии, где, в частности, даётся теория построения «номограмм из выравненных точек», а также указаны принципы построения номограмм для уравнений с бóльшим числом параметров.

В предыдущем параграфе было указано, что для решения уравнений типа

$$x^n + px + q = 0 \tag{4}$$

достаточно на чертеже с построенной кривой

$$y = x^n \tag{5}$$

построить прямую

$$y = -px - q \tag{6}$$

и определить абсциссы точек пересечения её с кривой (5).

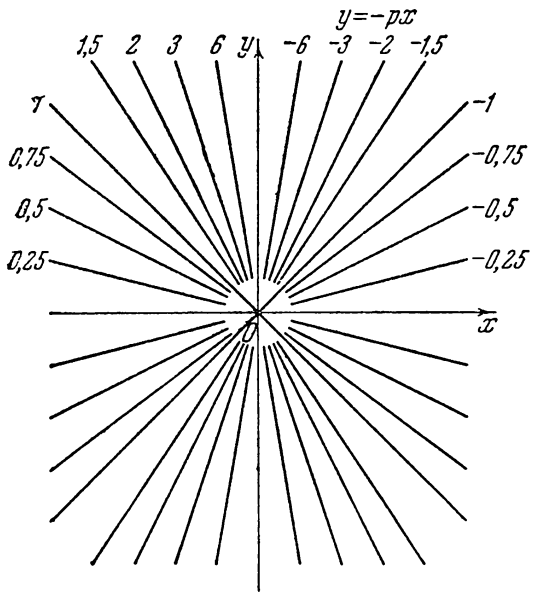

Рис. 13.

Но вместо построения прямых (6) можно воспользоваться транспарантом с начерченным на нём семейством прямых $y = -px$ при разных значениях p (рис. 13). Наложив такой транспарант на чертёж в соответствии с размерами начальной ординаты $-q$ прямой (6) и выбрав прямую на транспаранте с числовой отметкой, равной коэффициенту при x в данном уравнении, можно сразу найти корни уравнения (4).

Однако для построения номограммы уравнений разобранного типа можно использовать и совершенно другой принцип.

Рассмотрим, например, уравнение (1)

$$x^2 + px + q = 0.$$

Действительное число x_0 будет корнем уравнения (1) в том случае, если коэффициенты p и q этого уравнения связаны соотношением

$$q = -x_0 p - x_0^2. \tag{7}$$

Если будем по оси абсцисс откладывать значения p, а по оси ординат — значения q, то соотношению (7) будет соответствовать некоторая прямая (с угловым коэффициентом $-x_0$ и начальной ординатой $-x_0^2$).

Координаты любой точки этой прямой будут коэффициентами некоторого квадратного уравнения, одним из корней которого будет x_0.

Например, прямая MN: $q=\frac{1}{2}p-\frac{1}{4}$ на рис. 14 соответствует $x_0=-0{,}5$; поэтому для точек (1; 0,25), (— 1,5; — 1), лежащих на прямой MN, получим уравнения $x^2+x+0{,}25=0$, $x^2-1{,}5x-1=0$, для которых $x=-\frac{1}{2}$ будет корнем.

Начертив семейство прямых (7), соответствующих различным значениям x_0, получим так называемую сетчатую номограмму уравнения (1).

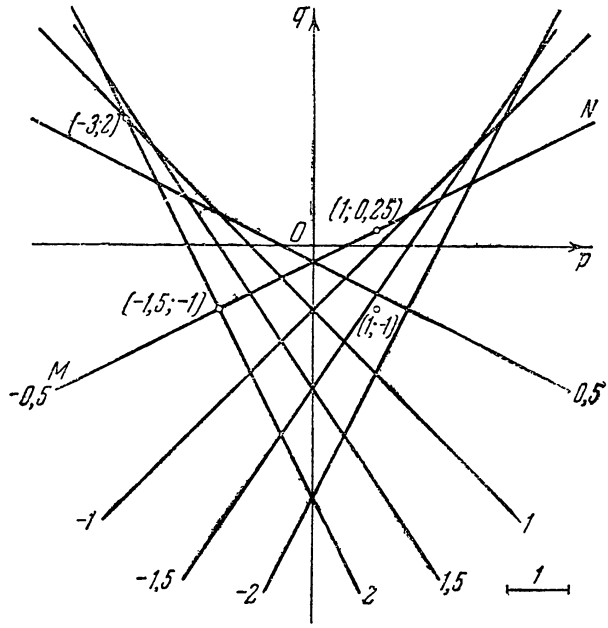

Рис. 14.

Для определения корней уравнения (1) при известных p и q достаточно на номограмме найти точку (p, q) и определить числовые отметки прямых, проходящих через эту точку; например, через точку (—3, 2) проходят прямые с отметками $x_1=1$ и $x_2=2$; следовательно, эти числа и будут корнями уравнения $x^2-3x+2=0$.

Найдём ещё корни уравнения $x^2+x-1=0$.

Хотя на рис. 14 нет прямых, проходяших через точку (1, — 1), но нетрудно на-глаз оценить, что числовые отметки прямых, которые при более частой сетке линий на номограмме прошли бы через точку (1, — 1), были бы: $x_1\approx 0{,}6$ и $x_2\approx -1{,}6$.

Конечно, точность решения будет тем больше, чем через меньшие интервалы изменяются на номограмме числовые отметки x.

Однако на рис. 14 имеются точки, через которые не может пройти ни одна из прямых номограммы; это значит, что корни соответствующих уравнений будут мнимыми. Легко проверить, что все точки такого рода лежат выше параболы $q=\frac{p^2}{4}$, т. е. для них $\frac{p^2}{4}-q<0$.

Номограмма, начерченная на рис. 14, называется *сетчатой*, так как на ней имеется по существу сетка из трёх [1]) семейств линий:

I. Рассмотренные прямые: $x = \text{const}$.

II. Прямые, параллельные оси абсцисс: $q = \text{const}$.

III. Прямые, параллельные оси ординат: $p = \text{const}$.

Недостаток сетчатых номограмм заключается в обилии линий, затрудняющем поиски решения.

Значительно удобнее так называемые номограммы из выравненных точек.

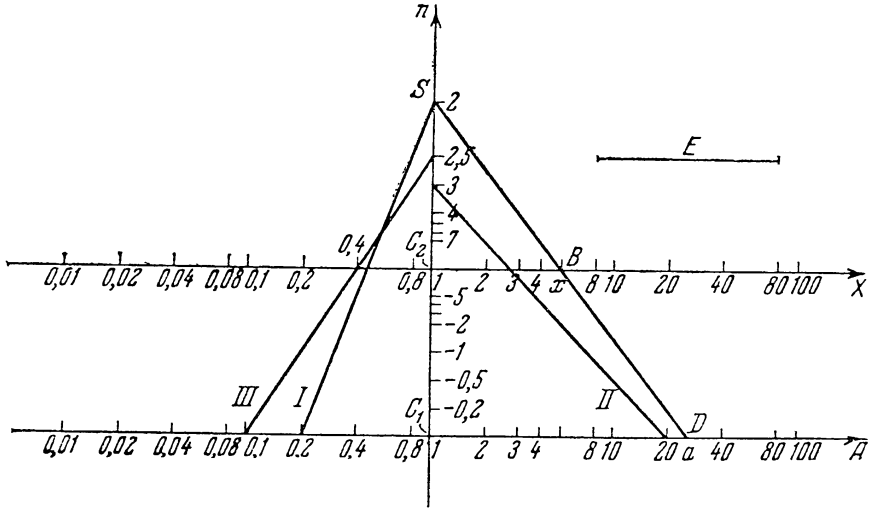

Рис. 15.

Рассмотрим уравнение

$$x^n - a = 0. \tag{3}$$

Если нас ингересуют лишь положительные корни уравнения (3) (при $a > 0$), то, логарифмируя (3), получим:

$$\lg x = \frac{1}{n} \lg a. \tag{3'}$$

Пусть на двух параллельных прямых C_1A и C_2X построены так называемые логарифмические шкалы, т. е. при выбранной единице масштаба E для расстояний C_1D и C_2B от начал шкал до точек с произвольными числовыми отметками a и x будем соответственно иметь (рис. 15):

$$C_1D = E \lg a,$$

$$C_2B = E \lg x.$$

[1]) На рис. 14 семейства линий II и III не нарисованы; если номограмма строится на клетчатой бумаге, то достаточно построить лишь семейство линий I.

Если на прямой C_1C_2 взять точку S, для которой $\frac{SC_1}{SC_2}=n$, то из подобных треугольников SC_1D и SC_2B будем иметь: $\frac{C_1D}{C_2B}=\frac{SC_1}{SC_2}$, т. е.

$$\frac{\lg a}{\lg x}=n. \tag{8}$$

Следовательно, при данных n и a достаточно найти соответствующие точки S и D на прямых C_1C_2 и C_1A номограммы и с помощью линейки найти точку B, лежащую с точками S и D на одной прямой (номограмма из выравненных точек!).

Числовая отметка точки B и даст корень уравнения (3).

Конечно, номограмму на рис. 15 можно использовать и для решения уравнения

$$b^z-a=0. \tag{9}$$

В этом случае точки с числовыми отметками a и b берутся на шкалах C_1A и C_2X, а искомое значение z ищется на прямой C_1C_2.

На рис. 15 показано для примера решение уравнений

I. $x^2-0{,}2=0$, положительный корень $x_1\approx 0{,}45$;

II. $x^3-20=0$, $x_1\approx 2{,}7$;

III. $0{,}4^z-0{,}1=0$, $z\approx 2{,}5$.

Опишем, наконец, построение номограмм из выравненных точек для уравнений вида (1), (2) и т. п.

На плоскости Oxy построим кривую, заданную уравнениями

$$\left.\begin{aligned} x&=\frac{1}{1+z},\\ y&=-\frac{zs}{1+z}\end{aligned}\right\} \tag{10}$$

(здесь z — параметр; если его исключить из данных уравнений, то получим: $y=-\frac{(1-x)^s}{x^{s-1}}$).

Придавая различные значения параметру z, нетрудно составить таблицу соответствующих друг другу значений x и y. Например, при $s=2$ будем иметь:

z	0	1	2	3	4		9		∞	$-0{,}5$	$-0{,}9$
x	1	0,5	$^1/_3$	0,25	0,2	...	0,1	...	0	2	10
y	0	$-0{,}5$	$-^4/_3$	$-2{,}25$	$-3{,}2$		$-8{,}1$		$-\infty$	$-0{,}5$	$-8{,}1$

z	$-0{,}99$	-1	$-1{,}1$	$-1{,}5$	-2	-3		-11		$-\infty$
x	100	$\pm\infty$	-10	-2	-1	$-0{,}5$	...	$-0{,}1$	...	0
y	$-88{,}01$	$\mp\infty$	12,1	4,5	4	4,5		12,1		$+\infty$

Если нанесём на чертёж точки, соответствующие найденным парам чисел x и y, проставляя около каждой точки соответствующее ей значение параметра z, то получим «криволинейную шкалу» для переменного z (при $s=2$ это будет гипербола, изображённая на рис. 16).

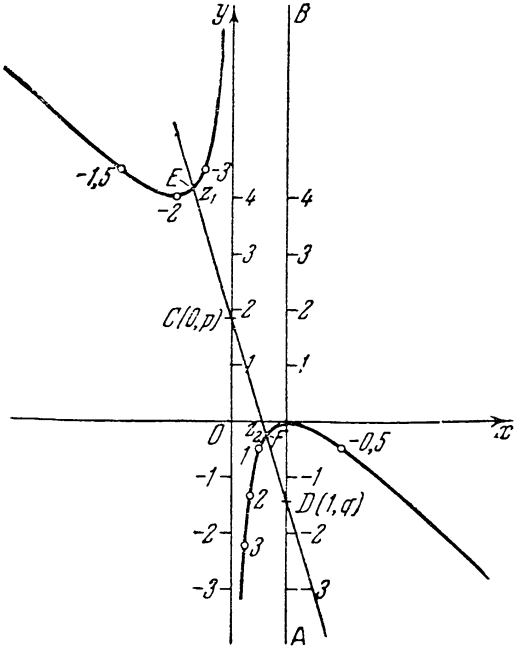

Рис. 16.

Кроме этой криволинейной шкалы, нанесём на оси Oy равномерную шкалу p, а на прямой AB — равномерную шкалу q.

Если при значениях p и q, равных коэффициентам данного нам уравнения

$$z^s+pz+q=0, \quad (11)$$

взять на чертеже точки $C(0, p)$ и $D(1, q)$, то прямая CD пересечёт, вообще говоря, построенную кривую в точках E и F с числовыми отметками z_1 и z_2 (а при $s>2$ могут быть и другие точки пересечения).

Применив известное условие принадлежности трёх точек одной прямой к точкам $C(0, p)$, $D(1, q)$ и $E\left(\frac{1}{1+z_1}, -\frac{z_1^s}{1+z_1}\right)$, получим:

$$\begin{vmatrix} 1 & 0 & p \\ 1 & 1 & q \\ 1 & \frac{1}{1+z_1} & -\frac{z_1^s}{1+z_1} \end{vmatrix} = 0$$

или $-\frac{z_1^s}{1+z_1}+\frac{p}{1+z_1}-p-\frac{q}{1+z_1}=0$, откуда

$$z_1^s+pz_1+q=0. \quad (12)$$

Соотношение (12) показывает, что числовая отметка точки E, лежащей с точками C и D на одной прямой, будет корнем уравнения (11). То же, конечно, можно сказать и о числовых отметках других точек пересечения прямой CD с криволинейной шкалой z.

Замечая, что положительные числовые отметки z расположены на части кривой, заключённой между прямыми Oy и AB, можно ограничиться лишь этой частью криволинейной шкалы, что позволит определить положительные корни уравнения (11); для определения же отрицательных корней достаточно в уравнении (11) сделать

подстановку $z=-t$ и искать положительные корни t_1, t_2, ... уравнения $(-t)^s - pt + q = 0$, т. е. при чётном s — уравнения

$$t^s - pt + q = 0, \quad (13)$$

а при нечётном s — уравнения

$$t^s + pt - q = 0. \quad (14)$$

Тогда $-t_1$, $-t_2$, ... будут искомыми отрицательными корнями уравнения (11).

На рис. 17 приведена номограмма для определения положительных корней уравнений (1) и (2), причём единица масштаба по оси Ox взята значительно больше, чем по оси Oy.

Там же показано для примера решение уравнений

$z^2 + 5z - 1{,}5 = 0$ ($z_1 \approx -5{,}3$, $z_2 \approx 0{,}3$);

$z^3 - 2z + 6 = 0$ ($z_1 \approx -2{,}2$; положительных корней нет).

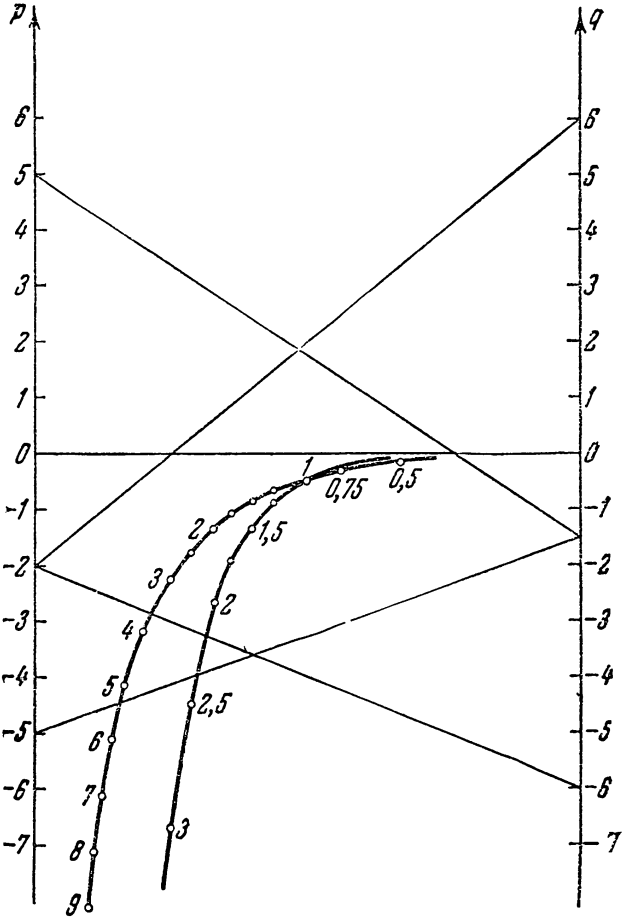

Рис. 17.

Конечно, на рис. 14, 15, 17 даны лишь схематические изображения номограмм. Рабочие номограммы чертятся бóльших размеров с бóльшим числом числовых отметок на шкалах или с более густой сеткой прямых, снабжённых числовыми отметками.

§ 16. Решение систем уравнений

Уже в § 11 при определении исходного приближения корней системы уравнений с двумя неизвестными был использован графический способ.

В § 14 оказалось целесообразным заменить уравнение четвёртой степени системой уравнений, допускающей простое графическое решение.

Если дана система уравнений

$$\Phi(x, y) = 0, \quad \Psi(x, y) = 0, \tag{1}$$

то для определения её действительных корней достаточно определить точки пересечения кривых, соответствующих этим уравнениям; поскольку точки будут лежать одновременно на обеих кривых, то их координаты (x_1, y_1), (x_2, y_2) и т. д. будут удовлетворять и тому и другому уравнениям, т. е. будут корнями системы (1).

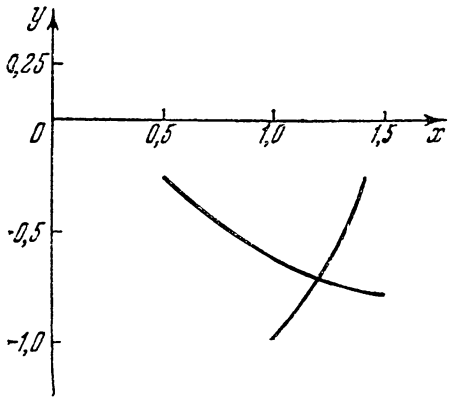

Рис. 18.

Построение кривых сильно упрощается, если уравнения могут быть решены относительно одного из неизвестных, или оба неизвестных можно выразить как функции одного и того же параметра (см., например, § 15 — построение рис. 16).

Практически удобно строить сначала кривые в мелком масштабе по сравнительно далеко разбросанным точкам, что даёт грубо приближённые значения корней; затем участки кривых, примыкающие к точкам пересечения, следует вычертить в крупном масштабе, что даст искомые корни с большей точностью. После этого можно или ещё раз повторить этот приём, или же уточнять полученные значения корней по способу Ньютона или по способу итерации.

Решим для примера систему

$$x - \cos y - 0{,}44 = 0, \quad y + \sin x - 0{,}23 = 0.$$

Грубый набросок кривых

$$y = 0{,}23 - \sin x, \quad x = 0{,}44 + \cos y$$

показывает, что корни следует искать в окрестности точки $(1; -0{,}5)$.

Чертим сначала кривые в мелком масштабе, для чего вычисляем координаты их точек [1]):

x	0,5	0,75	1	1,25	1,5
$y = 0{,}23 - \sin x$	−0,25	−0,45	−0,61	−0,72	−0,77
y	0	−0,25	−0,5	−0,75	−1
$x = 0{,}44 + \cos y$	1,44	1,41	1,32	1,17	0,98

[1]) При вычислении значений функций, в которых аргумент входит и самостоятельно и под знаком тригонометрических функций, удобно пользоваться таблицами тригонометрических функций с аргументом, выраженным в радианах (см., например, Н. М. Бескин, Вопросы тригонометрии и её преподавание, 1950, стр. 132—137.

Из рис. 18 видно, что кривые пересекаются вблизи точки (1,2; —0,7).

В окрестности этой точки составляем более подробные таблицы:

x	1,18	1,2	1,22
$y = 0{,}23 - \sin x$	0,6946	0,7020	0,7091

y	—0,68	—0,7	—0,72
$x = 0{,}44 + \cos y$	1,2176	1,2048	1,1918

Из рис. 19, начерченного в значительно более крупном масштабе, находим:

$$x_1 \approx 1{,}204, \quad y_1 \approx -0{,}704.$$

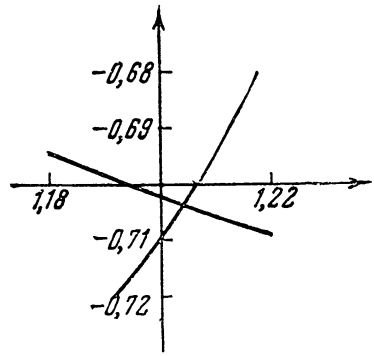

Рис. 19.

В общем случае построение кривой, соответствующей уравнению $F(x, y) = 0$, связано с утомительными вычислениями, так как для определения значений y, соответствующих определённому значению $x = x_0$, надо решить уравнение

$$F(x_0, y) = 0. \qquad (2)$$

Правда, иногда можно привлечь на помощь математический анализ: найти асимптоты кривой, точки перегиба и т. п., но полностью избежать необходимости решения уравнения (2), как правило, не удаётся.

Отметим, что особенно просто проводится решение системы линейных уравнений

$$\left.\begin{aligned} a_1 x + b_1 y &= c_1, \\ a_2 x + b_2 y &= c_2, \end{aligned}\right\} \qquad (3)$$

так как для каждого уравнения достаточно найти лишь две (не очень близкие!) точки и соединить их по линейке прямой линией. Конечно, при небольших, «круглых» коэффициентах нет нужды применять графический способ; но при громоздких коэффициентах (а такие случаи встречались нам и в § 11, когда мы искали поправки $\bar{\alpha}$ и $\bar{\beta}$ в способе Ньютона, и в § 12 при отыскании коэффициентов α, β, γ, δ в способе итерации) графический способ очень удобен.

Для случая трёх уравнений с тремя неизвестными

$$f(x, y, z) = 0, \tag{4}$$

$$\varphi(x, y, z) = 0, \tag{5}$$

$$\psi(x, y, z) = 0 \tag{6}$$

геометрическими образами, соответствующими этим уравнениям в прямоугольных декартовых координатах, будут поверхности, а корнями системы — координаты точек, общих всем трём поверхностям.

Простейшим по идее методом, позволяющим на чертеже находить точки пересечения трёх поверхностей, является изображение их в проекциях с числовыми отметками; например, придавая переменному z в уравнении (4) какое-нибудь конкретное значение $z = z_0$, получим соотношение между переменными x и y:

$$f(x, y, z_0) = 0, \tag{7}$$

которому на чертеже в плоскости Oxy будет соответствовать некоторая кривая, называемая *горизонталью поверхности* (4); около этой горизонтали ставят числовую отметку z_0.

Если для каждого из уравнений начертить на одном и том же чертеже семейства горизонталей, например, с отметками $z = 0$, 1, 2, 3, ..., -1, -2, ..., то, соединяя точки пересечения горизонталей (с одинаковыми числовыми отметками) поверхностей (4) и (5), получим изображение в проекциях с числовыми отметками линии их пересечения; затем так же ищется линия пересечения поверхностей (4) и (6); координаты общих точек этих двух линий дадут искомые корни (координата z будет равна числовой отметке точки пересечения линий, определяемой «на-глаз» по числовым отметкам соседних точек). Поясним описанный способ примером.

Пример. Решить систему уравнений

$$x - y + 3z - 7 = 0,$$

$$x^2 + y^2 - z^2 = 0,$$

$$x - y^2 + 2z - 2{,}5 = 0,$$

или

$$y = x + 3z - 7, \tag{8}$$

$$x^2 + y^2 = z^2, \tag{9}$$

$$x = y^2 - 2z + 2{,}5. \tag{10}$$

При $z = 1, 2, 3, 4, \ldots$ получим для этих уравнений соответственно семейства прямых, окружностей и парабол. На рис. 20 указаны линии пересечения (*8—10*), (*9—10*) поверхностей, попарно взятых [линия (*8—9*) не указана, но она должна пройти через точки *A*, *B*, *C*, *D*].

Эти линии проходят через точки A, B, C, D, координаты которых

$$\left.\begin{array}{llll} x_1 \approx 1{,}5, & x_2 \approx 1, & x_3 \approx -2{,}9, & x_4 \approx -4, \\ y_1 \approx 2{,}0, & y_2 \approx -1{,}3, & y_3 \approx -0{,}9, & y_4 \approx 1{,}3, \\ z_1 \approx 2{,}5, & z_2 \approx 1{,}6, & z_3 \approx 3, & z_4 \approx 4{,}1 \end{array}\right\} \quad (11)$$

и дают нам искомые корни.

Конечно, приближённые значения корней (11) можно графически же исправить, если начертить в более крупном масштабе участки поверхностей, прилегающие к точкам их пересечения.

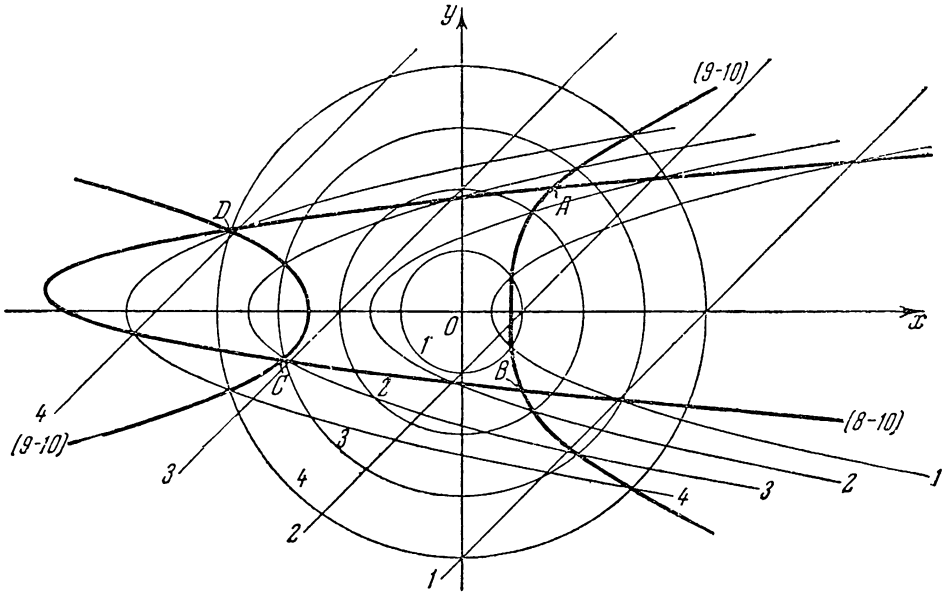

Рис. 20.

В общем случае вычерчивание горизонталей поверхностей представляет очень сложную задачу. Иногда помощь могут оказать специальные транспаранты; например, для рассмотренной системы полезна была бы «прозрачная» парабола $x = y^2$.

Особенно просто графически решаются системы линейных уравнений, так как для них семействами горизонталей будут семейства параллельных прямых, а прямые пересечения плоскостей и в проекциях с числовыми отметками изображаются прямыми, что делает очень простым их построение.

На рис. 21 показано графическое решение системы уравнений

$$\left.\begin{array}{r} 1{,}5x - 1{,}5y - 2{,}2z - 0{,}5 = 0, \\ 0{,}8x + 1{,}7y + z - 4{,}2 = 0, \\ 2{,}3x - 1{,}4y + 2z - 3{,}8 = 0. \end{array}\right\} \quad (12)$$

Приведя, например, третье уравнение к виду

$$y = \frac{2{,}3}{1{,}4}x + \frac{2z - 3{,}8}{1{,}4},$$

определяем при $z = 0$ и при двух каких-нибудь значениях x соответствующие им значения y; потом то же самое делаем при $z = 1$, что даёт две горизонтали третьей плоскости.

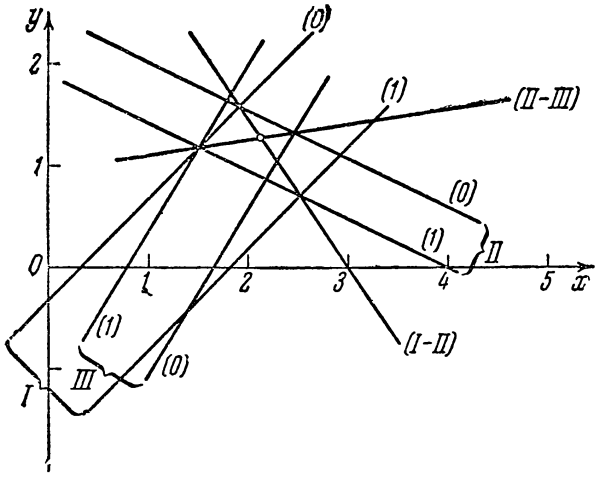

Рис. 21.

Вычисления удобно проводить на логарифмической линейке, так как большой точности, как правило, не требуется.

Из рис. 21 видим, что $x_1 \approx 2{,}1$ $y_1 \approx 1{,}3$ $z_1 \approx 0{,}3$. Отметим, что системы линейных уравнений и с бо́льшим числом неизвестных допускают простое графическое решение, основанное на других принципах [1]). Графическое же решение систем уравнений общего вида более чем с тремя неизвестными чрезвычайно сложно.

Задачи к главе IV

25. Сколько действительных корней имеет уравнение

$$10 \sin x - x = 0?$$

Найти графически наименьший положительный корень с точностью до 0,01.

26. Найти графически иррациональный корень уравнения $2^x - 4x = 0$ с точностью до 0,001.

27. Решить графически систему уравнений

$$2{,}41x - 2{,}15y = 0{,}32, \quad 0{,}45x - 1{,}14y = 1{,}6.$$

28. Показать, что подстановкой вида $x = \alpha z$ любое кубическое уравнение

$$a_0x^3 + a_1x^2 + a_2x + a_3 = 0$$

может быть приведено к виду

$$z^3 + z^2 + Az + B = 0.$$

Указать способ графического решения кубических уравнений, опирающийся на это преобразование.

---

[1]) См., например, Д. Н. Головнин, Графическая математика, 1931, стр. 130—134.

29. Проводя окружности на чертеже с параболой $y=x^2$, решить уравнения:

а) $x^4-14x^2+24x-8=0$, в) $x^4+2x^2+x+2=0$,

б) $x^4-7x^2-4x-20=0$, г) $x^3+5x-11=0$

(в последнем случае, умножив обе части уравнения на x, перейти к уравнению четвёртой степени с лишним корнем: $x=0$).

30. Найти по способу Лилля корни уравнений:

а) $x^3+x^2+x-7=0$, г) $3x^2+x-2=0$,

б) $x^4+x^3-x-1=0$, д) $x^2+4x+4=0$,

в) $x^3-2=0$, е) $x^2-4x+5=0$.

31. Найти графически с точностью до 0,01 наименьшие положительные корни системы

$$\left.\begin{aligned} x&=\operatorname{tg} y,\\ y&=\operatorname{tg} x. \end{aligned}\right\}$$

32. Построить в большом масштабе номограммы из выравненных точек уравнений (1), (2) из § 15 и применить их к решению уравнений:

а) $3x^2-11x-4=0$,

б) $5x^2+8x-13=0$,

в) $x^3-10x+2=0$.

33. Решить, проводя прямые на чертеже с начерченной параболой $y=x^2$ уравнения:

а) $x^2-0{,}05x+0{,}0007=0$,

б) $60x^2-x-0{,}01=0$

(предварительно сделать подстановки вида $x=\alpha z$).

34. Используя переход к более крупному масштабу, вычислить корни системы (12) из § 16 с точностью до 0,01.

35. Используя переход к более крупному масштабу, вычислить корни системы, состоящей из уравнений (8), (9), (10) из § 16, с точностью до 0,01.

ДОБАВЛЕНИЯ

1. Краткие исторические сведения

Сравнительно недавняя расшифровка клинописных табличек, памятников вавилонской культуры, показала, что уже в начале второго тысячелетия до нашей эры математические знания вавилонян были довольно обширны. Им была, например, известна формула для приближённого извлечения корней квадратных из чисел

$$\sqrt{a^2+b^2}\approx a+\frac{b}{2a} \quad (\text{при } b<a).$$

Они умели решать задачи, приводящие к квадратным и даже к простейшим кубическим уравнениям. Например, в одной из дошедших до нас задач дело сводится — в современных обозначениях — к решению уравнения

$$12x^3+x^2=1+\frac{45}{60}$$

(напомним, что у вавилонян была в употреблении так называемая шестидесятиричная система счисления).

После умножения обеих частей уравнения на 12^2 оно было приведено к виду

$$(12x)^3+(12x)^2=4\cdot 60+12,$$

откуда

$$12x=6.$$

Решение преобразованных к такому виду уравнений осуществлялось, вернее всего, с помощью специальных таблиц (также дошедших до нас), которые в наших обозначениях выглядят примерно так:

$$\text{если } z^3+z^2=2, \quad \text{то } z=1,$$

$$\text{если } z^3+z^2=12, \quad \text{то } z=2,$$

$$\dots\dots\dots\dots\dots\dots$$

$$\text{если } z^3+z^2=5\cdot 60^2+4\cdot 60+12, \quad \text{то } z=26,$$

$$\text{если } z^3+z^2=5\cdot 60^2+40\cdot 60+12, \quad \text{то } z=27$$

и т. д.

Египтянам, чья культура развивалась параллельно с вавилонской, решение уравнений второй и третьей степеней было, повидимому, неизвестно; правда, о математических знаниях египтян мы судим в основном по двум математическим папирусам: лондонскому (550×32 *см*^{2}) и московскому (550×8 *см*^{2}).

Некоторые из задач, помещённых в этих папирусах, приводят к уравнениям вида (в современных обозначениях)

$$ax + bx + cx + \ldots = d,$$

где a, b, c, ... — некоторые целые числа или дроби, числитель которых обычно равнялся единице.

Для решения таких уравнений применялся следующий приём, который впоследствии получил название *правила ложного положения* (regula falsi): вычисляли при каком-нибудь $x = x_1$ значение левой части уравнения

$$ax_1 + bx_1 + cx_1 + \ldots = d_1,$$

а затем взятое число x_1 делили на полученное число и умножали результат на d $\left(x = \frac{x_1 d}{d_1}\right)$.

Позже большую роль в разных странах играло так называемое *правило двух ложных положений* (по существу линейное интерполирование), описанное в § 7.

Индусы решали с его помощью задачи, которые в нашей трактовке приводили бы к уравнениям первой степени; для них это правило давало точное решение. Арабский математик ал-Кархи (XI в.) получил, повидимому, с помощью линейного интерполирования формулу

$$\sqrt{a^2 + r} \approx a + \frac{r}{2a + 1};$$

Леонардо Пизанский (в 1202 г.) аналогичным путём пришёл к формуле

$$\sqrt[3]{a^3 + r} \approx a + \frac{r}{3a^2 + 3a + 1}.$$

Кстати, отметим, что Леонардо, должно быть применяя способ линейного интерполирования, смог вычислить с большой точностью положительный корень уравнения $x^3 + 2x^2 + 10x = 20$: $x_1 = 1^\circ\ 22'\ 7''\ 42'''\ 33^{IV}\ 4^V\ 40^{VI}$ (в шестидесятиричной системе); погрешность найденного им значения корня была меньше чем 10^{-10}.

Сравнительно малое влияние на численное решение уравнений оказала греческая математика. Греческих геометров периода V—I вв. до н. э. вычислительная математика (логистика), как правило, мало интересовала. Даже в сочинениях, где встречалось извлечение корней квадратных (например, у Архимеда), нет точных указаний на то, как эта операция выполнялась. Впрочем, позже в «Метрике»

Герона александрийского (жившего около I—II в. н. э. или несколько раньше) указан способ извлечения корня квадратного, который в наших обозначениях может быть выражен так:

$$\sqrt{A} \approx \frac{1}{2}\left(a + \frac{A}{a}\right),$$

где a — приближённое значение корня с недостатком, а $\frac{A}{a}$ — с избытком.

Безусловно, грекам были известны и другие приёмы извлечения корней квадратных и даже кубических, но дошедшие до нас сведения позволяют делать лишь более или менее вероятные предположения как относительно сущности, так и относительно теоретической основы этих способов. Авторы некоторых работ по истории математики склонны считать, что греческие математики применяли для вывода приближённых формул линейное интерполирование, хотя весьма вероятно, что источники этих формул следует искать «в геометрической алгебре», столь свойственной греческой математике [1]).

Впрочем, у Птолемея (II в. н. э.) извлечение квадратных корней осуществляется уже в шестидесятиричной системе приблизительно так же, как делается в употребляемом теперь алгорифме.

Для геометрических задач третьей степени — см., например, уравнения (5), (6) введения — греки интересовались больше геометрическим построением отрезков, являвшихся решением задачи, чем вычислением их длин.

Применяемый в настоящее время алгорифм извлечения корней квадратных и кубических из чисел, дающий одну за другой цифры искомого корня, был уже в VI в. н. э. известен (конечно, внешне в другой форме) индусам, создавшим, как известно, позиционную десятичную систему записи чисел.

Переданный через арабов в Европу, способ этот в 1600 г. был приспособлен Вьета для вычисления корней алгебраических уравнений.

Однако лишь после опубликования (в 1819 г.) правила Горнера, позволяющего легко производить преобразования, встречающиеся в способе Вьета, способ этот приобрёл форму, описанную в § 4.

В конце XVII в. Ньютон предлагает свой метод численного решения уравнений, а Рефсон придаёт этому методу ту форму, в какой он описан в § 7.

К тому же времени относится зарождение способа итерации.

Исследования Лагранжа (XVII в.) по теории непрерывных дробей привели, в частности, к способу решения уравнений, описанному в § 5.

---

[1]) См., например С. Я. Лурье, Приближённые вычисления в Древней Греции, Архив истории науки и техники, вып. 4, 1934, стр. 21—47.

Открытый Н. И. Лобачевским [1]) способ решения алгебраических уравнений являлся крупным шагом вперёд, так как дал в руки вычислителей мощное средство для вычисления и действительных, и мнимых корней алгебраических уравнений.

Обилие технических расчётов, не требующих большой точности, привело в конце прошлого века к развитию графических методов решения уравнений и к созданию номографии.

В настоящее время внимание вычислителей направлено на развитие численных методов решения алгебраических уравнений специального вида (так называемых *вековых уравнений*), играющих большую роль в механике, астрономии, алгебре и т. д. Один из наиболее практичных методов решения «вековых» уравнений принадлежит акад. А. Н. Крылову.

2. Советы преподавателям и рекомендуемая литература

В настоящей статье не исчерпаны, конечно, все существующие способы численного и графического решения уравнений, да в этом и нет необходимости. Нам кажется, что, познакомившись со статьёй и проделав упражнения, читатель убедится в возможности вычисления корней любых уравнений и систем уравнений с любой степенью точности (при наличии соответствующих таблиц).

Следует, однако, обратить внимание преподавателей на два вопроса, не рассмотренные в настоящей статье: 1) решение уравнений с помощью построения графиков логарифмов функций, где очень полезным является применение так называемой логарифмической бумаги; 2) графическое решение систем линейных уравнений со многими неизвестными. Оба эти вопроса освещены в указанной ниже книге Головнина; они представляют хороший материал для школьных математических кружков.

Приведённые в статье списки задач без труда могут быть дополнены самим читателем, для чего можно посоветовать либо составление уравнений с заведомо известными корнями, например,

$$f(x)=(x^2-4x+5)(x^2-x-7)=x^4-5x^3+2x^2+23x-35=0$$

и т. п., либо решение случайно взятых уравнений или систем уравнений разными способами, что позволяет легко проверять правильность решения.

Считаем полезным обратить особое внимание преподавателей на графические способы решения уравнений, позволяющие, как было показано в § 14 и 16, находить действительные корни систем уравнений с одним и с двумя неизвестными с любой степенью точности.

Решение алгебраических и простейших трансцендентных уравнений табличным и графическим способами следует рекомендовать

[1]) См. подстрочное примечание на стр. 353.

ещё и потому, что это приучает к пользованию различными таблицами, помогает учащимся осознать роль радианного измерения углов, роль графического изображения функций и т. п.

Заметим, кстати, что при построении графиков и вообще при численных способах решения уравнений весьма полезными оказываются такие приборы, как счёты, логарифмическая линейка и особенно арифмометр; численное решение уравнений будет содействовать их популяризации в школе.

Как известно, в средней школе даётся геометрическая интерпретация решения простейших уравнений; важно, однако, создать у учащихся убеждение в том, что графически можно не только интерпретировать решение уравнений, но в некоторых случаях быстрее и проще находить корни уравнений, для чего, может быть, следует даже в классе решать уравнения вроде

$$x^2 - 1{,}63x - 2{,}47 = 0.$$

Графические методы решения уравнений особенно выгодно использовать в математических кружках; например, поручив кому-нибудь из кружковцев вычертить на миллиметровке кривые $y = x^2$, $y = x^3$; $y = \sin x$, $y = x^3 + x^2$ (см. задачу 28) и т. п. или номограммы (в большом масштабе) для уравнений (1), (2), (3) из § 15, можно затем на кружке провести упражнения на графическое решение уравнений.

Тем из читателей, кто захотел бы познакомиться с различными деталями, относящимися к численному и графическому решению уравнений, можно посоветовать прочесть следующие книги:

1. Безикович Я. С., Приближённые вычисления, изд. 6-е, Гостехиздат, Л. — М., 1949.

2. Глаголев А. А., Номография для школьников, 1935.

3. Глаголев Н. А., Теоретические основы номографии, изд. 2-е, ОНТИ, М. — Л., 1934.

4. Головнин Д. Н., Графическая математика, ГОНТИ, М. — Л., 1931.

5. Крылов А. Н., Лекции о приближённых вычислениях, изд. 5-е, Гостехиздат, М. — Л., 1950.

6. Млодзеевский Б. К., Решение численных уравнений, 1924.

7. Сушкевич А. К., Основы высшей алгебры, изд. 4-е, Гостехиздат, М. — Л., 1941.

8. Рунге К., Графические методы математических вычислений, перевод с нем., 1932.

9. Скарборо Дж., Численные методы математического анализа, перевод с англ., ГТТИ, М. — Л., 1934.

10. Уиттекер Э. и Робинсон Г., Математическая обработка результатов наблюдений, перевод с англ., ГОНТИ, М. — Л., 1935.

11. Runge, Praxis der Gleichungen, 1921.

12. Runge und König, Numerisches Rechnen, 1924.

13. Willers, Methoden der praktischen Analysis, 1928.

Кроме того, обратим внимание читателей на следующие таблицы, очень полезные при вычислениях.

14. Петерс И., Шестизначные таблицы тригонометрических функций, ГТТИ, М.—Л., 1932.

В этих таблицах аргумент изменяется через 10″.

15. Таблицы Барлоу квадратов, кубов, корней квадратных, корней кубичных и обратных величин целых чисел от 1 до 10000, изд. 2-е, ОНТИ НКТП, М.—Л., 1936.

16. Таблицы значений тригонометрических функций для аргумента, выраженного в радианах — в книге: Н. М. Бескин, Вопросы тригонометрии и её преподавание, 1950, стр. 132—137.

АЛФАВИТНЫЙ УКАЗАТЕЛЬ

Автоморфизм нормального поля 282
Алгебраическая дробь 129, 130, 244, 245, 254
— — , алгебраическая точка зрения 253
— — , функциональная точка зрения 251, 253, 254
Алгебраические дроби, действия над ними 245
— — , равенство 251, 252
Алгебраическое дополнение элемента определителя 33, 35—37
— уравнение, вычисление мнимых корней 350, 390
— — , разрешимость в радикалах 273, 275, 276, 279, 281
— — , условие разрещимости в квадратных радикалах 303
— — , — — в радикалах 232
— — , численные методы решения 317
Алгорифм деления с остатком 145, 147
— Евклида 148, 150, 269, 325
Ал-Кархи 379, 413
Архимеда задача 315

Базис конечного расширения 301
— пространства 13, 18, 20, 62, 64, 121
— — ортонормальный 88, 112
Билинейная форма 87
Бюдана-Фурье теорема 331

Вековое уравнение 415
Вектор 11, 19, 42, 43, 82, 84, 101
— , зеркальное отражение 103
— n-мерный 19
— собственный 116, 119, 120
Вектор, умножение на число 11, 14, 20, 48
Векторное пространство 42—45, 56, 58
— равенство, правила оперирования 46
Векторы, свойства операций над векторами 46—49
Высота члена многочлена 256
Высший член многочлена 256
Вьета 414
— формулы 208—210

Галуа группа уравнения 282, 285, 286
— поле 281, 285
— — уравнения 232
— теория 299, 300
Гаусс 201
Гаусса леммы 182, 183
Гельфонд А. О. 136
Гомоморфизм 293
Горизонталь поверхности 408, 409
Горнера способ вычисления действительного корня уравнения 332, 336
— — извлечения корня третьей степени 374
— схема 160—162, 320, 321
Граница верхняя отрицательных корней 324
— — положительных корней 318
— нижняя отрицательных корней 320
— — положительных корней 324
Границы корней уравнения, вычисление способом Маклорена 319, 320
— — — , — — Ньютона 319, 320
Греффе 344
Группа интранзитивная 298
Группа нормального поля 285
— подстановок n-й степени 272
— транзитивная 297, 298
— уравнения 282, 283, 285, 286
— — , свойства 290—292

Даламбер 201
Даламбера лемма 192
Данделен 344
Двучлен, условие неприводимости в поле 273
Декарта теорема 331
Деление многочлена на линейный двучлен 159—161

Деление многочленов 143
Делимость многочленов 142—145, 148
— целых чисел 142, 148
Делители нуля 140, 165
Делоне Б. Н. 299, 300
Дискриминант уравнения квадратного 209
— — третьей степени 214
Длина вектора 84, 85
Доморяд А. П. 365

Евклида алгорифм 148, 150, 269, 325

Жирар 201

Задача Архимеда 314
— об удвоении куба 307, 399
— о построении правильных многоугольников 399
— отделения корней 317, 324
— о трисекции угла 307, 314, 399
— уничтожения иррациональности в знаменателе 262, 265

Извлечение корня квадратного из действительного числа 372
— — — из комплексного числа 206
— — n-й степени из действительного числа 375
— — — — из комплексного числа 204
— — третьей степени из действительного числа 374
Изоморфизм 286, 293
— между кольцом многочленов и трансцендентным расширением 137—139
— тождественный 290
Инверсия 25
Интерполирование линейное 359
— параболическое 359
Интерполяционный многочлен 358

Кардан 219
Кардана формулы 211
Кольцо многочленов от двух неизвестных 235
— — от нескольких неизвестных 235, 237
— — от одного неизвестного 135, 137, 141
— с делителями нуля 140
— целых чисел 145
Координаты вектора 13,19
— — относительно базиса 62, 63
Корень алгебраического уравнения 163
— многочлена 162
— — k-кратный 163
Корень n-й степени из единицы 204
— первообразный n-й степени из единицы 204
Корни рациональные многочлена 169—173
Коэффициенты многочлена 131, 237
Крамера правило 41
Кратность корня многочлена 163
— неприводимого многочлена 159
Кронекера-Капелли теорема 70
Крылов А. Н. 415

Лагранжа способ вычисления действительного корня уравнения 336, 338, 341, 414
— — извлечения квадратных корней 372
Леверье 315
Леонардо Пизанский 379, 413
Лилля графический метод решения алгебраических уравнений 397
Линдеман 135
Линейная комбинация векторов 11—13, 49
Линейное преобразование (отображение), см. преобразование линейное
Линейность функции 17
Лиувилль 135
Лобачевский Н. И. 343, 415
Лобачевского способ вычисления мнимых корней 350, 355
— — решения алгебраических уравнений 343, 344, 415
— — , случай действительных корней, почти одинаковых по абсолютной величине 353
Логарифмическая шкала 402

Маклорена способ определения границы корней 319, 320
Матрица 44
— вырожденная 98
— единичная 98
— квадратная 94, 96
— линейного преобразования 106, 109
— обратная 98, 99
— ортогональная 90, 99, 112
— перехода 64, 89
— , произведение на число 44
— симметричная 112
— системы 70
— — расширенная 70
— суммы и произведения линейных преобразований 107—110
— транспонированная 92
Метрические задачи 87
Минор матрицы 69

Минор определителя 34, 35
Многочлен, алгебраическая точка зрения 130, 167, 244
— , его значение при значениях неизвестных 141, 242
— , единственность представления в нормальном виде 131, 132, 239
— над областью целостности, свойство 141
— неприводимый в поле 155
— — , свойства 156, 157
— нулевой степени 145, 155
— однородный 240
— от нескольких неизвестных 236
— от одного неизвестного 131, 134
— приводимый в поле 155
— , признак неприводимости Эйзенштейна 185
— примитивный 182
— противоположный 135
— , разложение на неприводимые многочлены 156—159
— , — по степеням $x - c$ 321
— с действительными коэффициентами 200, 201
— симметрический 254
— с комплексными коэффициентами 199, 200
— с рациональными коэффициентами, разложение на неприводимые многочлены 174, 184, 185
— с целыми коэффициентами, условия отсутствия целых и рациональных корней 173
— , условие равенства нулю 132, 237, 242
— , функциональная точка зрения 129, 130, 142, 167, 244
— четвёртой степени, условие неприводимости в поле рациональных чисел 178
Многочлены взаимно простые 150, 153, 154
— над полем рациональных чисел 168
— — — , теория делимости 143, 148, 155
— равные 131, 132, 239, 243
— , совпадающие с точностью до множителя нулевой степени 144
Муавра формула 205

Наибольший общий делитель многочленов 148—151
Неизвестное 131, 139
Непрерывность функции в области комплексных чисел 188
Номограмма 399
— из выравненных точек 402, 403
— сетчатая 401, 402
Нормальный вид многочлена 237
Нулевой вектор 49, 57
Ньютон 378, 414
Ньютона способ обобщённый решения трансцендентных уравнений 363, 366, 367, 375, 377
— — определения границы корней 319, 320
— — решения систем уравнений 380, 383
— — — трансцендентных уравнений 357, 359, 360, 362, 371, 372, 375, 379, 414

Область рациональности уравнения 232
— целостности 141
— — бесконечная 142, 167, 189
Образ вектора 101, 110
— элемента множества 100
Общий делитель многочленов 148, 151
Определитель второго порядка 17, 18
— , вычисление определителя 34, 37
— матрицы 96
— произвольного порядка 21—24, 27—29, 31
— , разложение по элементам ряда 34, 36
— , свойства определителя 17, 18, 21—24, 29—32, 37
— системы уравнений 38
— третьего порядка 24
Остаток 145
— от деления многочлена на линейный двучлен 160, 162
Отображение векторного пространства 101
— гомоморфное 293
— линейное плоскости 103
— — пространства 103
— множеств 100, 101

Пересечение подпространств 57
Перестановка 24—26
— нечётная 25
— чётная 25
Подкольцо 135
Подполе 226
Подпоследовательность 197
Подпространство векторного пространства 56, 57
— нульмерное 76
— , порождаемое множеством векторов 58

Подпространство собственное 58
Подстановка единичная (тождественная) 271
— n-й степени 269
— обратная 272
Подходящая дробь 338—340
Поле 141
— алгебраических дробей 246, 247, 251
— действительных чисел 302
— комплексных чисел 188, 189, 302
— нормальное (поле Галуа) 232, 281, 285, 286
— числовое 19, 188
Последовательность комплексных чисел 195
— — — ограниченная 196, 197
— — — расходящаяся 195
— — — сходящаяся 195
— — — , условие сходимости 195
Правило двух ложных положений 357, 413
— Крамера 41
Правило ложного положения 357, 413
Предел числовой последовательности 195
Преобразование алгебраического уравнения 205
— координат 88, 90
— — ортогональное 90, 91
— линейное 101, 106, 109, 112
— — вырожденное 110, 111
— — , геометрические свойства 110
— — невырожденное 110—112, 122
— — ортогональное 112—114, 119
— — , представление матрицами 105
— — , — произведением ортогонального и симметрического 122
— — , примеры 102
— — , произведение на сопряжённое 122
— — симметрическое 119, 120
— — , условие ортогональности 113, 119
— — , условия невырожденности 111
— обратное 112
— плоскости в себя 121
— подобия 116
— положительно определённое 121
— пространства 101
— симметрическое 112, 119
— — , геометрическая характеристика 114
— — плоскости 115—117
— — трёхмерного пространства 117, 119
— , собственное значение 116
— сопряжённое 117—119
Преобразование сопряжённое произведения преобразований 123
— тождественное 111, 112
— , условие сопряжённости 118
Присоединение элемента к полю 226
Проблема решения алгебраических уравнений в радикалах 232
— резольвент 300
Продолжение изоморфизма 286—288
Произведение автоморфизмов 283
— алгебраических дробей 245
— матриц 93, 95, 99
— многочленов 132, 134, 237
— отображений 104, 105
— подстановок 270
— примитивных многочленов 182
— функций 168
Пространство n-мерное 20, 42, 43
Птолемей 414

Равенство алгебраических дробей 251, 252
— векторов 11, 13
— многочленов 131, 132, 139, 167, 239, 243
Радикал 202
Размерность векторного пространства 56, 58
— подпространства решений однородной системы 73
— пространства многочленов 59
Разность векторов 12
— многочленов 135, 237
Разрешающее уравнение (резольвента) 267
Ранг матрицы 69, 82
— — , способ вычисления 81
— системы векторов 55, 56, 66
Растяжение пространства 121
Расширение кольца 135—137
— конечное 301, 302
— поля 226
— — алгебраическое 225, 302
— — простое 226
— — — алгебраическое 226
— — — трансцендентное 226
— — путём присоединения чисел 230, 231
Резольвента 267, 299
— кубическая 176, 177
Результант системы уравнений 81
Рефсон 414
Решение алгебраических уравнений в радикалах 232
— — — , графический метод Лилля 393

Решение алгебраических уравнений, способ Горнера 332, 336
— — — , — Лагранжа 336, 338, 341, 414
— — — , — Лобачевского 343, 344, 350, 353, 355, 415
— двучленных уравнений 372
— квадратного уравнения 206
— — — , графический метод 392, 394, 399
— — — с помощью номограмм 400, 401
— систем линейных уравнений, графический метод 407, 409
— — уравнений, графический метод 405, 406, 408
— — — , способ итерации 383, 388, 389
— — — , — Ньютона 380, 383
— трансцендентных уравнений, обобщение способа Ньютона 363, 366, 367, 375, 377
— — — , способ итерации 367, 369, 371, 372, 378
— — — , — линейного интерполирования 357—359, 360, 362, 413
— — — , — Ньютона 357, 359, 360, 362, 371, 372, 375, 379, 414
— уравнений в радикалах 202
— — с помощью номограмм 399
— — , табличный способ 315, 330, 331, 357, 362, 363
— уравнения третьей степени 211, 219
— — — — , графический метод 395, 399
— — четвёртой степени 221, 222
— — — — , графический метод 395, 396
Руффини-Абеля теорема 279, 281

Сдвиг 103
Симметрическая алгебраическая дробь 254, 261
— группа n-й степени 272
— функция 254
Симметрические многочлены, приложения 262, 265—267
Симметрический многочлен 254
— — , выражение через основные симметрические многочлены 258, 259, 261
— — неоднородный 261
— — основной (элементарный) 255
— — , основные свойства 255
Симметрическое преобразование, см. преобразование симметрическое
Система векторов 64
— — линейно зависимая 49
— — — независимая 49, 51, 68
— — максимальная линейно независимая 55
— — , условие линейной зависимости 50
— — , — — независимости 68
— линейных неоднородных уравнений, совокупность решений 76
— — — — , условие единственности решения 61, 62
— — — — , — существование решения 60, 61, 70—73
— — однородных уравнений, совокупность решений 73
— — — — , условие существования решения 72, 73
— — уравнений 14, 18, 38, 39, 60, 70
— — — , основная теорема 40
— — — с двумя неизвестными 78
— — — с тремя неизвестными, геометрическое место решений 78
— однородных уравнений, соответствующая данной системе 76
— уравнений, существование решений в различных числовых полях 83
— функций Штурма 325, 329
— элементов линейно зависимая 301
— — — независимая 301
Системы векторов эквивалентные 51
— уравнений эквивалентные 81
Скалярное произведение векторов 84—88, 114
Сложение алгебраических дробей 245
— линейных преобразований 103
— матриц 100
— многочленов 132—134, 237
— рациональных функций 253
Старший коэффициент многочлена 131
— член многочлена 131
Степень конечного расширения 301
— многочлена от нескольких неизвестных 239, 240
— — от одного неизвестного 139, 141
— члена многочлена от нескольких неизвестных 240
Сумма алгебраических дробей 245
— векторов 11, 13, 48
— матриц 44
Сумма многочленов 132, 237
— n-мерных векторов 20
— отображений 104
— функций 168

Сумма элементов векторного пространства 42
Схема многочлена 397

Тейлора формула для многочленов 321
Теорема Бюдана-Фурье 331
— Декарта 331
— Кронекера-Капелли 70
— о делении с остатком 145
— о единственности выражения симметрических многочленов через основные симметрические многочлены 261
— — — поля алгебраических дробей 246
— о замене 51, 53, 54
— о непрерывности многочлена 189—191
— о неразрешимости общих алгебраических уравнений выше четвёртой степени в радикалах 279, 281
— о разложении многочлена на неприводимые многочлены 156—159
— о разрешимости алгебраического уравнения в квадратных радикалах 303
— основная алгебры 188, 191, 192, 199
— — — , следствия 199
— — о неразрешимости алгебраических уравнений с симметрической группой в радикалах 297
— — о симметрических многочленах 255, 257
— — о совместности систем уравнений 40
— о степени произведения многочленов 141, 241, 242
Теорема о существовании поля алгебраических дробей 247
— — — трансцендентного элемента 136
— Штурма 325—327, 329
Теоремы о конечных расширениях 302
Транспозиция 25, 298
Транспонирование, действие на сумму и произведение матриц 95
— матриц 92
— таблицы определителя 32

Угол между двумя векторами 84
Умножение автоморфизмов 283
— алгебраических дробей 245
— матриц 93, 94, 96, 100
— многочленов 132—134, 237
— отображений 104
Умножение подстановок 270, 271
— рациональных функций 253
Уравнение двучленное 202, 204
— квадратное с комплексными коэффициентами 208, 209
— третьей степени 210, 215
— — — каноническое 210
— — — , неприводимый случай 217
— — — с действительными коэффициентами, исследование 216, 217, 219
— — — , условия разрешимости в квадратных радикалах 304, 305
— четвёртой степени 220, 267
— — — , условия разрешимости в квадратных радикалах 307, 308
Уравнения выше четвёртой степени, неразрешимость в радикалах 273
— с симметрической группой 294

Фаддеев Д. К. 300
Феррари 220
Форма m-й степени 240
Функции Штурма 325, 327—329
Функция 100
— линейная 17
— непрерывная в области комплексных чисел 188, 189
— — комплексного переменного, свойства 196
— рациональная 252

Центральная кривая второго порядка, упрощение уравнения 124

Частное 145
Чеботарёв Н. Г. 299
Число алгебраическое относительно поля 226
— действительных корней многочлена с действительными коэффициентами 201
— — — уравнения 325, 331
— измерений 56
— комплексных корней многочлена 200
— корней многочлена 164, 165
— отрицательных корней уравнения 327, 331
— положительных корней уравнения 327, 331
— трансцендентное 135, 315
— — относительно поля 226
Член многочлена с наибольшей степенью 256
Члены многочлена 131, 134, 237

Шатуновский С. О. 299
Штурма система функций 325, 329
— теорема 325—327, 329
— функции 325, 327—329

Эйзенштейна признак неприводимости многочленов 185
Эйлер Л. 222
Эйлера способ решения уравнений четвёртой степени 222
Эквивалентность систем векторов, свойства 51
— — линейных уравнений 81
Элемент кольца трансцендентный 135, 136
Элементы поля сопряжённые 290
Эрмит 135